aws

Guia do Desenvolvedor

AWS Lake Formation

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

AWS Lake Formation: Guia do Desenvolvedor

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

As marcas comerciais e imagens de marcas da Amazon não podem ser usadas no contexto de nenhum produto ou serviço que não seja da Amazon, nem de qualquer maneira que possa gerar confusão entre os clientes ou que deprecie ou desprestigie a Amazon. Todas as outras marcas comerciais que não pertencem à Amazon pertencem a seus respectivos proprietários, que podem ou não ser afiliados, patrocinados pela Amazon ou ter conexão com ela.

Table of Contents

O que AWS Lake Formationé 1
Características do Lake Formation 2
Ingestão e gerenciamento de dados 2
Gerenciamento de segurança 3
Trazer seus dados para o Catálogo de Dados 5
Como funciona 6
Fluxo de trabalho de gerenciamento de permissões do Lake Formation 7
Permissões de metadados 8
Gerenciar o acesso às do armazenamento 11
Cross-account compartilhamento de dados em Lake Formation 13
Componentes do Lake Formation 14
Lake Formation 14
API e interface de linha de comando do Lake Formation 14
Outros serviços da AWS 14
Terminologia do Lake Formation 15
Data lake 15
Acesso aos dados 15
Modo de acesso híbrido 15
Blueprint 16
Fluxo de trabalho 16
catálogo de dados 16
Dados subjacentes 17
Entidade principal 17
Administrador do data lake 17
AWS integrações de serviços com Lake Formation 17
Recursos adicionais do Lake Formation 19
Blogs 19
Palestras técnicas e webinars 20
Arquitetura moderna 20
Recursos de data mesh 20
Guias de práticas recomendadas 20
Introdução ao Lake Formation 21
Introdução 22
Inicial completa AWS tarefas de configuração 22

Inscreva-se em um Conta da AWS ... 22
Configurar AWS Lake Formation ... 22
Configure os recursos de formação de lagos usando CloudFormation modelo ... 23
Crie um administrador de data lake ... 24
Alterar o modelo de permissão padrão ou usar o modo de acesso híbrido ... 29
Atribuir permissões aos usuários do Lake Formation ... 30
Como configurar um local no Amazon S3 para o data lake ... 32
(Opcional) Configurações externas de filtragem de dados ... 33
(Opcional) Conceder acesso à chave de criptografia do catálogo de dados ... 34
(Opcional) Criar um perfil do IAM para fluxos de trabalho ... 34
Atualizar as permissões de dados AWS Glue para o modelo do Lake Formation ... 36
Sobre permissões padrão ... 36
Listar permissões existentes ... 38
Configurar permissões do Lake Formation ... 40
Dar permissões do IAM aos usuários ... 41
Mudar para o modelo de permissões do Lake Formation ... 41
Etapa 5: proteja os novos recursos do catálogo de dados ... 44
Etapa 6: fornecer aos usuários uma nova política do IAM ... 45
Etapa 7: limpar políticas do IAM existentes ... 46
Configurando endpoints da VPC da Amazon (AWS PrivateLink) ... 47
Considerações sobre endpoints da VPC do Lake Formation ... 47
Criar um endpoint da VPC de interface para o Lake Formation ... 48
Criar uma política de endpoint da VPC para o Lake Formation ... 48
Tutoriais ... 50
Criando um data lake a partir de um AWS CloudTrail origem ... 51
Público-alvo ... 52
Pré-requisitos ... 53
Etapa 1: Criar um usuário de analista de dados ... 54
Etapa 2: adicionar permissões para leitura AWS CloudTrail registros para a função de fluxo de trabalho ... 55
Etapa 3: Criar um bucket do Amazon S3 para o data lake ... 55
Etapa 4: Registrar um caminho do Amazon S3 ... 56
Etapa 5: Conceder permissões de local de dados ... 56
Etapa 6: Criar um banco de dados no catálogo de dados ... 57
Etapa 7: Conceder permissões de dados ... 57
Etapa 8: Usar um esquema para criar um fluxo de trabalho ... 58

Etapa 9: Executar o fluxo de trabalho 59
Etapa 10: Conceder SELECT nas tabelas 60
Etapa 11: consulte o data lake usando Amazon Athena 60
Criação de um data lake a partir de uma fonte JDBC 61
Público-alvo 62
Pré-requisitos 63
Etapa 1: Criar um usuário de analista de dados 63
Etapa 2: criar uma conexão no AWS Glue 64
Etapa 3: Criar um bucket do Amazon S3 para o data lake 65
Etapa 4: Registrar um caminho do Amazon S3 65
Etapa 5: Conceder permissões de local de dados 66
Etapa 6: Criar um banco de dados no catálogo de dados 66
Etapa 7: Conceder permissões de dados 67
Etapa 8: Usar um esquema para criar um fluxo de trabalho 67
Etapa 9: Executar o fluxo de trabalho 69
Etapa 10: Conceder SELECT nas tabelas 70
Etapa 11: consulte o data lake usando Amazon Athena 70
Etapa 12: Consulte os dados no data lake usando o Amazon Redshift Spectrum 71
Etapa 13: Conceder ou revogar permissões do Lake Formation usando o Amazon Redshift Spectrum 75
Como configurar permissões para formatos de tabela aberta no Lake Formation 76
Público-alvo 76
Pré-requisitos 77
Etapa 1: Provisionar os recursos 78
Etapa 2: Configurar permissões para uma tabela do Iceberg 80
Etapa 3: Configurar permissões para uma tabela do Hudi 86
Etapa 4: Configurar permissões para uma tabela do Delta Lake 88
Etapa 5: limpar AWS recursos 91
Como gerenciar um data lake usando o controle de acesso baseado em tags 91
Público-alvo 93
Pré-requisitos 94
Etapa 1: Provisionar os recursos 94
Etapa 2: registrar sua localização de dados, criar uma ontologia da tag do LF e conceder permissões 95
Etapa 3: Criar bancos de dados do Lake Formation 99
Etapa 4: Conceder permissões de dados 109

Etapa 5: Executar uma consulta no Amazon Athena para verificar as permissões 111
Etapa 6: limpar AWS os recursos 112
Como proteger os data lakes com controle de acesso em nível de linha 113
Público-alvo 113
Pré-requisitos 114
Etapa 1: Provisionar os recursos 115
Etapa 2: Consulta sem filtros de dados 116
Etapa 3: Configurar filtros de dados e conceder permissões 118
Etapa 4: Consulta com filtros de dados 120
Etapa 5: limpar AWS os recursos 122
Compartilhe seus dados com segurança usando o Lake Formation 122
Público-alvo 123
Definir as configurações do Lake Formation 124
Etapa 1: provisionar seus recursos usando AWS CloudFormation modelos 127
Etapa 2: Pré-requisitos de compartilhamento entre contas do Lake Formation 129
Etapa 3: Implementar o compartilhamento entre contas usando o método de controle de acesso baseado em tags 131
Etapa 4: Implementar o método de recurso nomeado 138
Etapa 5: limpar AWS os recursos 141
Compartilhamento de recursos do Catálogo de Dados com recursos externos Contas da AWS usando controle de acesso refinado 142
Público-alvo 143
Pré-requisitos 144
Etapa 1: Forneça acesso refinado a outra conta 145
Etapa 2: Forneça acesso refinado a um usuário na mesma conta 146
Permissões de integração ao Lake Formation 148
Visão geral das permissões do Lake Formation 149
Métodos para controle de acesso de alta granularidade 151
Controle de acesso a metadados 154
Controle de acesso a dados subjacente 158
Referência de personas e permissões do IAM do Lake Formation 164
AWS Lake Formation personas 164
AWS políticas gerenciadas para Lake Formation 166
Permissões sugeridas por personas 175
Alterando as configurações padrão do seu data lake 186
Permissões implícitas do Lake Formation 189

Referência de permissões do Lake Formation 191
Permissões do Lake Formation por tipo de recurso 191
Concessão e revogação da Lake Formation AWS CLI Comandos da do 194
Permissões do Lake Formation 199
Integrar o Centro de Identidade do IAM 214
Pré-requisitos para integrar o Centro de Identidade do IAM ao Lake Formation 215
Conectar o Lake Formation ao Centro de Identidade do IAM 219
Atualizar integração com o Centro de Identidade do IAM Identity 223
Excluir uma conexão do Lake Formation com o Centro de Identidade do IAM 225
Conceder permissões a usuários e grupos 226
Incluindo o contexto do usuário do IAM Identity Center nos CloudTrail registros 229
Adicionar uma localização do Amazon S3 ao seu data lake 231
Requisitos para funções usadas para registrar locais 232
Registrando uma localização do Amazon S3 240
Registrando uma localização criptografada do Amazon S3 244
Registrando uma localização do Amazon S3 em outra AWS account 249
Registrando uma localização criptografada do Amazon S3 em AWS contas 251
Cancelar o registro de uma localização do Amazon S3 256
Acessando locais do Amazon S3 257
Modo de acesso híbrido 261
Casos de uso comuns do modo de acesso híbrido 263
Como funciona o modo de acesso híbrido 265
Configurando o modo de acesso híbrido - cenários comuns 266
Removendo entidades principais e recursos do modo de acesso híbrido 287
Visualizando entidades principais e recursos no modo de acesso híbrido 288
Recursos adicionais do 289
Criação de objetos no AWS Glue Data Catalog 289
Criar um catálogo 290
Criação de um banco de dados 291
Criar tabelas 292
Criar visualizações do Catálogo de Dados 300
Importação de dados usando fluxos de trabalho 349
Esquema e fluxos de trabalho 350
Criação de um fluxo de trabalho 351
Executar um fluxo de trabalho 355
Trazer seus dados para o Catálogo de Dados 357

Federação de catálogos 359
Características da Federação de Catálogos 359
Limitações 360
Federar para Snowflake 360
Federar para Databricks 368
Trazer dados do Amazon Redshift para o Catálogo de Dados do 377
Benefícios principais 379
Perfis e responsabilidades 380
Pré-requisitos 381
Criar catálogos federados do Amazon Redshift 387
Visualizar objetos do catálogo 396
Atualizar um catálogo federado 398
Acessar um catálogo federado compartilhado 400
Excluir um catálogo federado 405
Consultar catálogos federados 406
Recursos adicionais do 407
Federar em fontes de dados externas 407
Fluxo de trabalho 408
Pré-requisitos 409
Criar um catálogo federado 413
Visualizar objetos do catálogo 419
Excluir um catálogo federado 421
Consultar catálogos federados 422
Recursos adicionais do 423
Integração de tabelas do Amazon S3 com o Data Catalog e o Lake Formation 423
Como funciona a integração do Data Catalog e do Lake Formation 424
Pré-requisitos 426
Habilitar a integração de Tabelas do Amazon S3 429
Criar bancos de dados e tabelas 434
Registrando um bucket de tabela do Amazon S3 em outra conta AWS 437
Conceder permissões 440
Alterando os controles de acesso para integração com tabelas do S3 443
Habilite o Lake Formation com a integração do S3 Tables 444
Mudança de Lake Formation para IAM 446
Criar um catálogo gerenciado do Amazon Redshift 449

Gerenciamento de permissões para dados em uma unidade de compartilhamento de dados do Amazon Redshift. ... 453
Pré-requisitos ... 455
Configuração de permissões para unidades de compartilhamento de dados do Amazon Redshift ... 456
Consultar bancos de dados federados ... 460
Gerenciamento de permissões em conjuntos de dados que usam repositórios de dados externos ... 461
Fluxo de trabalho ... 464
Pré-requisitos ... 464
Conectando o catálogo de dados a um repositório externo do Hive ... 467
Recursos adicionais do ... 470
Gerenciando permissões do Lake Formation ... 471
Conceder permissões de localização de dados ... 471
Concessão de permissões de localização de dados (mesma conta) ... 472
Concessão de permissões de localização de dados (conta externa) ... 475
Conceder permissões em um local de dados compartilhado com sua conta ... 478
Conceder permissões de dados ... 479
Permissões do IAM necessárias para conceder permissões do Lake Formation ... 481
Usar o método de recursos nomeados ... 484
Tag-based controle de acesso ... 510
Concedendo permissões de data lake usando o método LF-TBAC ... 571
Attribute-based controle de acesso ... 578
Exemplo de cenário de permissões ... 582
Filtragem de dados e segurança por célula ... 584
Filtros de data ... 586
Suporte PartiQL em expressões de filtro de linha ... 590
Permissões necessárias para consultar tabelas com filtragem em nível de célula ... 592
Como gerenciar filtros de dados ... 593
Visualizar permissões de banco de dados e tabelas ... 608
Revogar permissões usando o console ... 612
Cross-account compartilhamento de dados ... 613
Pré-requisitos ... 616
Como atualizar as configurações da versão de compartilhamento de dados entre contas ... 621
Compartilhando tabelas e bancos de dados do Catálogo de Dados entre Contas da AWS ou diretores do IAM de contas externas ... 627

Conceder permissões em um banco de dados ou tabela compartilhada com sua conta 633
Como conceder permissões de links de recursos 635
Como acessar os dados subjacentes de uma tabela compartilhada 638
Cross-account CloudTrail registro 640
Gerenciando permissões entre contas usando ambos AWS Glue e Lake Formation 645
Visualizando todas as concessões entre contas usando a operação de GetResourceShares API 647
Acessar e visualizar tabelas e bancos de dados compartilhados do catálogo de dados 649
Aceitando um convite AWS RAM de compartilhamento de recursos 650
Visualizando tabelas e bancos de dados compartilhados do catálogo de dados 652
Criação de links de recursos 654
Como funcionam os links de recursos 655
Como criar um link de recurso para uma tabela compartilhada 657
Como criar um link de recurso para um banco de dados compartilhado 660
Manipulação de links de recursos em AWS Glue APIs 664
Acessar tabelas entre regiões 668
Fluxos de trabalho 670
Como configurar o acesso à tabela entre regiões 674
Segurança 677
Proteção de dados 677
Criptografia em repouso 678
Segurança da infraestrutura 679
Prevenção contra o ataque do “substituto confuso” em todos os serviços 680
Registro de eventos de segurança em AWS Lake Formation 681
Integração com o Lake Formation 682
Como usar a integração de aplicativos Lake Formation 682
Como funciona a integração de aplicações do Lake Formation 683
Perfis e responsabilidades na integração de aplicativos do Lake Formation 685
Lake Formation fluxo de trabalho para operações de API de integração de aplicativos 686
Como registrar um mecanismo de consulta de terceiros 688
Como habilitar permissões para que um mecanismo de consulta de terceiros chame operações de API de integração de aplicativos 690
Integração de aplicativos para acesso total à tabela 694
Trabalhando com outros AWS serviços 697
Amazon Athena 700
Suporte a formatos de tabelas transacionais 702

Recursos adicionais do ... 705
Amazon Redshift Spectrum ... 705
Suporte para tipos de tabelas transacionais ... 706
Recursos adicionais do ... 708
AWS Glue ... 708
Suporte para tipos de tabelas transacionais ... 709
Recursos adicionais do ... 710
Amazon EMR ... 710
Suporte a formatos de tabelas transacionais ... 711
Recursos adicionais do ... 712
Quick ... 713
Recursos adicionais do ... 713
AWS CloudTrail Lago ... 713
Registrando chamadas da API AWS Lake Formation usando AWS CloudTrail ... 715
Informações sobre Lake Formation em CloudTrail ... 715
Noções básicas de eventos do Lake Formation ... 716
Rastreando o acesso aos dados CloudTrail com o Lake Formation ... 718
Práticas recomendadas, considerações e limitações do Lake Formation ... 721
Práticas recomendadas e considerações sobre compartilhamento de dados entre contas ... 721
Limitações de perfis vinculadas a serviços ... 724
Limitações de acesso aos dados entre regiões ... 725
Considerações e limitações das visualizações do catálogo de dados ... 725
Limitações de filtragem de dados ... 727
Notas e restrições para filtragem em nível de coluna ... 727
Limitações de filtragem no nível de célula ... 729
Considerações e limitações do modo de acesso híbrido ... 730
Limitações para trazer dados do armazém de dados do Amazon Redshift para o AWS Glue Data Catalog ... 732
Limitações da integração do catálogo do S3 Tables ... 734
Considerações e limitações do compartilhamento de dados de armazenamento de metadados do Hive ... 735
Limitações do compartilhamento de dados do Amazon Redshift ... 736
Limitações da integração com o Centro de Identidade do IAM ... 738
Considerações e práticas recomendadas de controle de acesso com base em tags do Lake Formation ... 739
Considerações sobre controle de acesso por atributo, limitações e regiões aceitas ... 742

Solução de problemas do Lake Formation 743
Solução de problemas gerais 743
Erro: permissões insuficientes do Lake Formation em <Amazon S3 location> 743
Erro: “permissões de chave de criptografia insuficientes para a API Glue” 744
Minha consulta Amazon Athena ou do Amazon Redshift que usa manifestos está falhando 744
Erro: "permissão(ões) do Lake Formation insuficiente(s): necessária a criação de tag no catálogo" 744
Erro ao excluir administradores de data lake inválidos 744
Erro: “O chamador está temporariamente não autorizado a realizar esta operação no recurso...” 744
Resolução de problemas de acesso entre contas 744
Eu concedi uma permissão para várias contas do Lake Formation, mas o destinatário não consegue ver o recurso 745
As entidades principais da conta do destinatário podem ver o recurso do catálogo de dados, mas não podem acessar os dados subjacentes 746
Erro: “Falha na associação porque o chamador não foi autorizado” ao aceitar um convite de compartilhamento AWS RAM de recursos 746
Erro: “não autorizado a conceder permissões para o recurso” 747
Erro: “Acesso negado para recuperar informações AWS da organização” 747
Erro: “organização <organization-ID> não encontrada” 747
Erro: "permissões insuficientes do Lake Formation: combinação ilegal" 747
ConcurrentModificationException em grant/revoke solicitações para contas externas 747
Erro ao usar o Amazon EMR para acessar dados compartilhados por meio de várias contas 748
Solução de problemas em esquemas e fluxos de trabalho 749
<role-ARN>Meu plano falhou com “Usuário: <user-ARN>não está autorizado a executar: iam: PassRole no recurso:” 749
Meu fluxo de trabalho falhou com “Usuário: <user-ARN>não está autorizado a executar: iam: PassRole no recurso:<role-ARN>” 750
Um crawler no meu fluxo de trabalho falhou com “o recurso não existe ou o solicitante não está autorizado a acessar as permissões solicitadas” 750
Um rastreador no meu fluxo de trabalho falhou com “Ocorreu um erro (AccessDeniedException) ao chamar a CreateTable operação...” 750
Problemas conhecidos do AWS Lake Formation 750
Limitação na filtragem de metadados da tabela 751

Problema ao renomear uma coluna excluída 752
Problema com a exclusão de colunas em tabelas CSV 752
As partições da tabela devem ser adicionadas em um caminho comum 752
Problema com a criação de um banco de dados durante a criação do fluxo de trabalho 752
Problema com a exclusão e a recriação de um usuário 753
As operações da API do catálogo de dados não atualizam o valor do parâmetro `IsRegisteredWithLakeFormation` 753
As operações do Lake Formation não oferecem suporte ao AWS Glue Schema Registry 753
Mensagem de erro atualizada 754
Lake FormationAPI do 755
Permissões 756
— operações — 756
— tipos de dados — 756
Configurações do data lake 757
— operações — 757
— tipos de dados — 757
Integração com o Centro de Identidade do IAM 757
— operações — 757
— tipos de dados — 757
Modo de acesso híbrido 758
— operações — 758
— tipos de dados — 756
Fornecimento de credenciais 758
— operações — 758
— tipos de dados — 759
Tags 759
— operações — 759
— tipos de dados — 760
APIs de filtro de dados 760
— operações — 760
— tipos de dados — 760
Tipos de dados comuns 761
ErrorDetail 761
Padrões de string 761
Regiões aceitas 763
Disponibilidade geral 763

AWS GovCloud (US) 763
Otimização de transações e armazenamento 763
Histórico do documento 766
Glossário da AWS 781
..... dcclxxxii

O que AWS Lake Formationé

Bem-vindo ao Guia do AWS Lake Formation desenvolvedor.

AWS Lake Formation ajuda você a governar, proteger e compartilhar dados de forma centralizada e global para análise e aprendizado de máquina. Com o Lake Formation, você pode gerenciar um controle de acesso refinado para seus dados de data lake no Amazon Simple Storage Service (Amazon S3) e seus metadados no AWS Glue Data Catalog.

O Lake Formation fornece seu próprio modelo de permissões que amplia o modelo de permissões do IAM. O modelo de permissões do Lake Formation permite acesso refinado aos dados armazenados em data lakes, bem como fontes de dados externas, como data warehouses do Amazon Redshift, bancos de dados Amazon DynamoDB e fontes de dados de terceiros por meio de um mecanismo simples de concessão ou revogação, muito parecido com um sistema de gerenciamento de banco de dados relacional (RDBMS). As permissões do Lake Formation são aplicadas usando controles granulares nos níveis de coluna, linha e célula em todos os serviços de AWS análise e aprendizado de máquina, incluindo Amazon Athena, Amazon Amazon Quick Redshift Spectrum, Amazon EMR e. AWS Glue

Com o modo de acesso híbrido Lake Formation para AWS Glue Data Catalog (Catálogo de dados), você pode proteger e acessar os dados catalogados usando as permissões do Lake Formation e as políticas de permissões do IAM para Amazon S3 AWS Glue e ações. Com o modo de acesso híbrido, os administradores de dados podem integrar as permissões do Lake Formation de forma seletiva e incremental, concentrando-se em um caso de uso do data lake por vez.

O Lake Formation também permite que você compartilhe dados interna e externamente em várias AWS organizações ou diretamente com diretores do IAM em outra conta Contas da AWS, fornecendo acesso refinado aos metadados do Catálogo de Dados e aos dados subjacentes.

Tópicos

- Características do Lake Formation
- AWS Lake Formation: como funciona
- Componentes do Lake Formation
- Terminologia do Lake Formation
- AWS integrações de serviços com Lake Formation
- Recursos adicionais do Lake Formation
- Introdução ao Lake Formation

Características do Lake Formation

O Lake Formation o ajuda a desfazer silos de dados e combinar diferentes tipos de dados estruturados e não estruturados em um repositório centralizado. Primeiro, identifique os armazenamentos de dados existentes no Amazon S3 ou nos bancos de dados relacionais e NoSQL e mova os dados para o seu data lake. Em seguida, rastreie, catalogue e prepare os dados para análise. Em seguida, forneça aos usuários acesso seguro de autoatendimento aos dados por meio de serviços de análise de sua escolha.

É possível usar o console do Lake Formation para criar catálogos federados de vários níveis no Data Catalog e unificar dados entre data lakes do Amazon S3 e data warehouses do Amazon Redshift. Você também pode integrar dados de seus bancos de dados operacionais Amazon DynamoDB, como fontes de dados de terceiros, como Google BigQuery, MySQL, entre outros. O Catálogo de Dados oferece um repositório centralizado de metadados que facilita o gerenciamento e a descoberta de dados em sistemas diferentes.

Para obter mais informações, consulte Trazendo seus dados para o AWS Glue Data Catalog.

Tópicos

- Ingestão e gerenciamento de dados
- Gerenciamento de segurança
- Trazer seus dados para o Catálogo de Dados

Ingestão e gerenciamento de dados

Importe dados de bancos de dados já existentes AWS

Depois de especificar onde estão seus bancos de dados existentes e fornecer suas credenciais de acesso, o Lake Formation lê os dados e seus metadados (esquema) para entender o conteúdo da fonte de dados. Em seguida, ele importa os dados para seu novo data lake e registra os metadados em um catálogo central. Com o Lake Formation, você pode importar dados de bancos de dados MySQL, PostgreSQL, SQL Server, MariaDB e Oracle executados no Amazon RDS ou hospedados no Amazon EC2. Tanto o carregamento de dados em massa quanto o incremental são suportados.

Importar dados de outras fontes externas

Você pode usar o Lake Formation para mover dados de bancos de dados on-premises conectando-se ao Java Database Connectivity (JDBC). Identifique suas fontes de destino e forneça credenciais

de acesso no console, e o Lake Formation lê e carrega seus dados no data lake. Para importar dados de bancos de dados diferentes dos listados acima, você pode criar trabalhos ETL personalizados com o. AWS Glue

Catalogue e rotule seus dados

Você pode usar AWS Glue rastreadores para ler seus dados no Amazon S3, extrair o esquema do banco de dados e da tabela e armazenar esses dados em um catálogo de dados pesquisável. Em seguida, use o Lake Formation Controle de acesso baseado em tags do Lake Formation (TBAC) para gerenciar as permissões em bancos de dados, tabelas e colunas. Para obter mais informações sobre como adicionar tabelas ao catálogo de dados, consulte Criação de objetos no AWS Glue Data Catalog.

Gerenciamento de segurança

Defina e gerencie controles de acesso

O Lake Formation fornece um único local para gerenciar os controles de acesso aos dados em seu data lake. Você pode definir políticas de segurança que restrinjam o acesso aos dados nos níveis de banco de dados, tabela, coluna, linha e célula. Essas políticas se aplicam a usuários e funções do IAM e a usuários e grupos durante a federação por meio de um provedor de identidade externo. Você pode usar controles refinados para acessar dados protegidos pelo Lake Formation no Amazon Redshift Spectrum, AWS Glue Athena, ETL e Amazon EMR para Apache Spark. Sempre que você criar identidades do IAM, siga as práticas recomendadas do IAM. Para obter mais informações, consulte Práticas recomendadas de segurança no Guia do usuário do IAM.

Modo de acesso híbrido

O modo de acesso híbrido do Lake Formation oferece a flexibilidade de habilitar seletivamente as permissões do Lake Formation para bancos de dados e tabelas no Data Catalog. Com o modo de acesso híbrido, agora você tem um caminho incremental que permite definir permissões do Lake Formation para um conjunto específico de usuários sem interromper as políticas de permissão de outros usuários ou workload existentes. Para obter mais informações, consulte Modo de acesso híbrido.

Implementar o registro em log de auditoria

O Lake Formation fornece registros de auditoria abrangentes CloudTrail para monitorar o acesso e mostrar a conformidade com políticas definidas centralmente. Você pode auditar o histórico de

acesso aos dados em serviços de análise e machine learning que leem os dados em seu data lake por meio do Lake Formation. Isso permite que você veja quais usuários ou funções tentaram acessar quais dados, com quais serviços e quando. Você pode acessar os registros de auditoria da mesma forma que acessa qualquer outro CloudTrail registro usando o console CloudTrail APIs e. Para obter mais informações sobre CloudTrail registros, consulteRegistrando chamadas da API AWS Lake Formation usando AWS CloudTrail.

Segurança por linha e célula

O Lake Formation fornece filtros de dados que permitem restringir o acesso a uma combinação de colunas e linhas. Use a segurança por linha e célula para proteger dados confidenciais, como Informações de Identificação Pessoal (PII). Para obter mais informações sobre segurança por linha, consulte Filtragem de dados e segurança por célula no Lake Formation.

Controle de acesso com base em tags

Use o controle de acesso por atributo do Lake Formation para gerenciar centenas ou até milhares de permissões de dados criando rótulos personalizados chamados tags LF. Agora você pode definir tags do LF e anexá-las a bancos de dados, tabelas ou colunas. Em seguida, compartilhe o acesso controlado por meio de serviços de análises, de machine learning (ML) e de extração, transformação e carregamento (ETL) para consumo. As tags do LF garantem que a governança de dados possa ser escalada facilmente substituindo as definições de políticas de milhares de recursos por algumas tags lógicas. O Lake Formation fornece uma pesquisa baseada em texto sobre esses metadados, para que seus usuários possam encontrar rapidamente os dados que precisam analisar.

Controle de acesso por atributo

Use o controle de acesso por atributo para conceder acesso aos objetos do Data Catalog. O controle de acesso baseado em atributos (ABAC) é uma estratégia de autorização que define permissões com base em atributos. AWS chama esses atributos de tags. É possível usar o ABAC para conceder acesso às entidades principais na mesma conta ou em outra conta nos recursos do Data Catalog. Qualquer entidade principal do IAM com chaves e valores correspondentes de tag do IAM ou de tag de sessão recebe acesso ao recurso. É necessário ter permissões para concessão nos recursos para fazer essas concessões.

Acesso entre contas

Os recursos de gerenciamento de permissões do Lake Formation simplificam a proteção e o gerenciamento de lagos de dados distribuídos em várias AWS contas por meio de uma abordagem centralizada, fornecendo controle de acesso refinado ao catálogo de dados e aos locais do Amazon

S3. Para obter mais informações, consulte Cross-account compartilhamento de dados em Lake Formation.

Trazer seus dados para o Catálogo de Dados

O recurso de federação permite criar catálogos federados e configurar permissões em conjuntos de dados armazenados em diferentes fontes de dados, como o Amazon Redshift, sem migrar dados ou metadados para o Amazon S3 ou o AWS Glue Data Catalog. É possível usar os seguintes métodos para trazer dados e gerenciar permissões em conjuntos de dados externos no Lake Formation:

Para acessar mais informações, consulte Trazer seus dados para o AWS Glue Data Catalog.

- Trazer dados nos data warehouses do Amazon Redshift para o AWS Glue Data Catalog: registre um namespace existente do Amazon Redshift ou um cluster no Data Catalog e crie um catálogo federado de vários níveis no Data Catalog.

  Você pode acessar seus dados usando qualquer mecanismo de consulta compatível com a especificação OpenAPI do catálogo REST do Apache Iceberg, como Amazon EMR Sem Servidor e Amazon Athena.

  Para obter mais informações, consulte Trazendo dados do Amazon Redshift para o AWS Glue Data Catalog.
- Federação no Catálogo de Dados a partir de fontes de dados externas — Conecte o Catálogo de Dados a fontes de dados externas usando AWS Glue conexões e crie catálogos federados para gerenciar centralmente as permissões de acesso em conjuntos de dados usando o Lake Formation. Não é necessária nenhuma migração de metadados para o Data Catalog.

  Para obter mais informações, consulte Federando em fontes de dados externas no AWS Glue Data Catalog.
- Integração dos buckets de tabela do Amazon S3 com o catálogo de dados — Você pode publicar e catalogar tabelas do Amazon S3 como objetos do catálogo de dados e registrar o catálogo como um local de dados do Lake Formation a partir do console do Lake Formation ou usando. AWS Glue APIs

  Para obter mais informações, consulte Integração de tabelas do Amazon S3 com AWS Glue Data Catalog and AWS Lake Formation.
- Criar catálogos para gerenciar Tabelas do Amazon Redshift no Data Catalog: você pode não ter um cluster produtor do Amazon Redshift ou uma unidade de compartilhamento de dados do Amazon Redshift disponível atualmente, mas deseja criar e gerenciar Tabelas do Amazon Redshift

usando o Data Catalog. Você pode começar criando um catálogo AWS Glue gerenciado usando a `glue:CreateCatalog` API ou o AWS Lake Formation console definindo o tipo de catálogo como `Managed` e `Catalog source` como Redshift.

Para obter mais informações, consulte Criação de um catálogo gerenciado do Amazon Redshift no AWS Glue Data Catalog.

- Integração do Lake Formation com o compartilhamento de dados do Amazon Redshift – Use o Lake Formation para gerenciar centralmente as permissões de acesso por banco de dados, tabela, coluna e linha das unidades de compartilhamento de dados do Amazon Redshift e restringir o acesso dos usuários a objetos em uma unidade de compartilhamento de dados.
- Conectando o catálogo de dados a metástores externos — Conecte-se AWS Glue Data Catalog a metastores externos para gerenciar permissões de acesso em conjuntos de dados no Amazon S3 usando o Lake Formation. Não é necessária nenhuma migração de metadados para o Data Catalog.

  Para obter mais informações, consulte Gerenciamento de permissões em conjuntos de dados que usam repositórios de dados externos.
- Integrando o Lake Formation com o AWS Data Exchange — O Lake Formation oferece suporte ao licenciamento de acesso aos seus dados por meio de. AWS Data Exchange Se você estiver interessado em licenciar seus dados do Lake Formation, consulte O que é o AWS Data Exchange no Guia do usuário do AWS Data Exchange .

AWS Lake Formation: como funciona

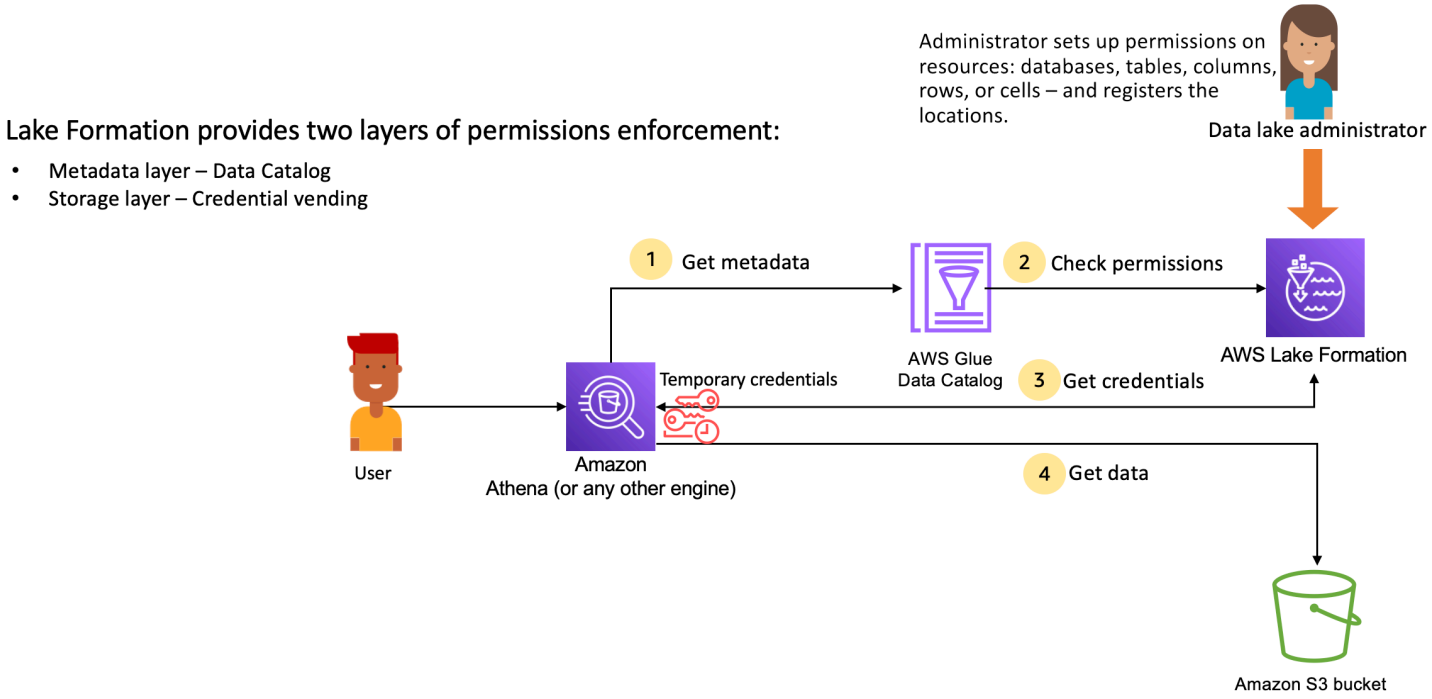
AWS Lake Formation fornece um modelo de permissões do sistema de gerenciamento de banco de dados relacional (RDBMS) para conceder ou revogar o acesso aos recursos do catálogo de dados, como bancos de dados, tabelas e colunas com dados subjacentes no Amazon S3. As permissões fáceis de gerenciar do Lake Formation substituem as políticas complexas de bucket do Amazon S3 e as políticas correspondentes do IAM.

No Lake Formation, você pode implementar permissões em dois níveis:

- Aplicação de permissões em nível de metadados nos recursos do catálogo de dados, como bancos de dados e tabelas
- Gerenciar permissões de acesso ao armazenamento nos dados subjacentes armazenados no Amazon S3 em nome de mecanismos integrados

Fluxo de trabalho de gerenciamento de permissões do Lake Formation

O Lake Formation se integra aos mecanismos analíticos para consultar armazenamentos de dados e objetos de metadados do Amazon S3 registrados no Lake Formation. O diagrama a seguir mostra como o gerenciamento de permissões funciona no Lake Formation.



Etapas de alto nível do gerenciamento de permissões do Lake Formation

Antes que o Lake Formation possa fornecer controles de acesso aos dados em seu data lake, um administrador de data lake ou um usuário com permissões administrativas configura políticas de usuário individuais da tabela do catálogo de dados para permitir ou negar acesso às tabelas do catálogo de dados usando as permissões do Lake Formation.

Em seguida, o administrador do data lake ou um usuário delegado pelo administrador concede permissões do Lake Formation aos usuários nos bancos de dados e tabelas do catálogo de dados e registra a localização da tabela no Amazon S3 no Lake Formation.

1. Obtenha metadados — Um principal (usuário) envia uma consulta ou um script de ETL para um mecanismo analítico integrado, como Amazon Athena, Amazon EMR ou AWS Glue Amazon Redshift Spectrum. O mecanismo analítico integrado identifica a tabela que está sendo solicitada e envia uma solicitação de metadados para o catálogo de dados.

2. Verificar permissões — O catálogo de dados verifica as permissões do usuário com o Lake Formation e, se o usuário estiver autorizado a acessar a tabela, retorna os metadados que o usuário tem permissão para ver para o mecanismo.
3. Obter credenciais — O catálogo de dados permite que o mecanismo saiba se a tabela é gerenciada pelo Lake Formation ou não. Se os dados subjacentes forem registrados no Lake Formation, o mecanismo analítico solicitará que o Lake Formation forneça acesso aos dados concedendo acesso temporário.
4. Obter dados — Se o usuário estiver autorizado a acessar a tabela, o Lake Formation fornecerá acesso temporário ao mecanismo analítico integrado. Ao usar o acesso temporário, o mecanismo analítico busca os dados do Amazon S3 e executa a filtragem necessária, como filtragem por coluna, linha ou célula. Quando o mecanismo termina de executar o trabalho, ele retorna os resultados para o usuário. Esse processo chamado de fornecimento de credenciais.

   Se a tabela não for gerenciada pelo Lake Formation, a segunda chamada do mecanismo analítico será feita diretamente para o Amazon S3. A política de bucket do Amazon S3 e a política de usuário do IAM em questão são avaliadas quanto ao acesso aos dados.

   Sempre que você usar as políticas do IAM, siga as práticas recomendadas do IAM. Para obter mais informações, consulte Práticas recomendadas de segurança no IAM no Guia do usuário do IAM.

Tópicos

- Permissões de metadados
- Gerenciar o acesso às do armazenamento
- Cross-account compartilhamento de dados em Lake Formation

Permissões de metadados

O Lake Formation fornece autorização e controle de acesso ao catálogo de dados. Quando um perfil do IAM faz uma chamada de API do catálogo de dados de qualquer sistema, o catálogo de dados verifica as permissões de dados do usuário e retorna somente os metadados que o usuário tem permissão para acessar. Por exemplo, se um perfil do IAM tiver acesso a somente uma tabela em um banco de dados e um serviço ou um usuário supondo que a função execute a operação `GetTables`, a resposta conterá somente uma tabela, independentemente do número de tabelas no banco de dados.

Configurações padrão - permissões de grupo **`IAMAllowedPrincipal`**

AWS Lake Formation, por padrão, define permissões para todos os bancos de dados e tabelas para um grupo virtual chamado `IAMAllowedPrincipal`. Esse grupo é único e visível somente dentro do Lake Formation. O `IAMAllowedPrincipal` grupo inclui todos os diretores do IAM que têm acesso aos recursos do catálogo de dados por meio das políticas principais e políticas de AWS Glue recursos do IAM. Se essas permissões existirem em um banco de dados ou tabela, todas as entidades principais terão acesso ao banco de dados ou à tabela.

Se você quiser fornecer permissões mais granulares em um banco de dados ou tabela, remova a permissão `IAMAllowedPrincipal` e o Lake Formation aplicará todas as outras políticas associadas a esse banco de dados ou tabela. Por exemplo, se houver uma política que permita ao Usuário A acessar o Banco de Dados A com permissões `DESCRIBE` e `IAMAllowedPrincipal` existir com todas as permissões, o Usuário A continuará executando todas as outras ações até que a permissão `IAMAllowedPrincipal` seja revogada.

Além disso, por padrão, o grupo `IAMAllowedPrincipal` tem permissões em todos os novos bancos de dados e tabelas quando eles são criados. Há duas configurações que controlam esse comportamento. O primeiro está na conta e isso permite Region-level isso para bancos de dados recém-criados, e o segundo está no nível do banco de dados. Para modificar a configuração padrão, consulte Alterar o modelo de permissão padrão ou usar o modo de acesso híbrido.

Conceder permissões

Os administradores do data lake podem conceder permissões do catálogo de dados às entidades principais para que elas possam criar e gerenciar bancos de dados e tabelas e acessar os dados subjacentes.

Permissões em nível de banco de dados e tabela

Quando você concede permissões no Lake Formation, o concedente deve especificar a entidade principal à qual conceder permissões, os recursos para os quais conceder permissões e as ações que o beneficiário deve ter acesso para realizar. Para a maioria dos recursos do Lake Formation, a lista de entidades principais e os recursos para conceder permissões são semelhantes, mas as ações que um beneficiário pode realizar diferem com base no tipo de recurso. Por exemplo, as permissões `SELECT` estão disponíveis para que as tabelas leiam as tabelas, mas as permissões `SELECT` não são permitidas nos bancos de dados. A permissão `CREATE_TABLE` é permitida em bancos de dados, mas não em tabelas.

Você pode conceder AWS Lake Formation permissões usando dois métodos:

- Método de recurso nomeado — Permite escolher nomes de banco de dados e tabelas enquanto concede permissões aos usuários.
- LF-Tag controle de acesso baseado (LF-TBAC) — Os usuários criam LF-Tags, os associam aos recursos do Catálogo de Dados, concedem `Describe` permissão LF-Tags, associam permissões a usuários individuais e escrevem políticas de permissões LF usando LF-Tags para usuários diferentes. Essas LF-Tag-based políticas se aplicam a todos os recursos do Catálogo de Dados associados a esses LF-Tag valores.

  > **Note**
  >
  > LF-Tags são exclusivos do Lake Formation. Elas só são visíveis no Lake Formation e não devem ser confundidas com tags AWS de recursos.

  LF-TBAC é um recurso que permite aos usuários agrupar recursos em categorias definidas pelo usuário LF-Tags e aplicar permissões a esses grupos de recursos. Portanto, é a melhor maneira de escalar as permissões em um grande número de recursos do catálogo de dados.

  Para obter mais informações, consulte Controle de acesso baseado em tags do Lake Formation.

Quando você concede permissões a uma entidade principal, o Lake Formation avalia as permissões como uma união de todas as políticas desse usuário. Por exemplo, se você tiver duas políticas em uma tabela para uma principal em que uma política concede permissões às colunas col1, col2 e col3 por meio do método de recurso nomeado, e a outra política concede permissões para a mesma tabela e principal para col5 e col6 a, as permissões efetivas serão uma união das permissões que seriam col1, col2, col3, col5 e col6. LF-Tags Isso também inclui filtros de dados e linhas.

Permissões de localização de dados

As permissões de localização de dados fornecem aos usuários não administrativos a capacidade de criar bancos de dados e tabelas em locais específicos do Amazon S3. Se um usuário tentar criar um banco de dados ou uma tabela em um local que não tenha permissão para criar, a tarefa de criação falhará. Isso evita que os usuários criem tabelas em locais arbitrários dentro do data lake e fornece controle sobre onde esses usuários podem ler e gravar dados. Há uma permissão implícita ao criar tabelas na localização do Amazon S3 dentro do banco de dados em que elas estão sendo criadas. Para obter mais informações, consulte Conceder permissões de localização de dados.

Crie permissões de tabela e banco de dados

Usuários não administrativos, por padrão, não têm permissões para criar bancos de dados ou tabelas em um banco de dados. A criação do banco de dados é controlada no nível da conta usando as configurações do Lake Formation para que somente entidades principais autorizadas possam criar bancos de dados. Para obter mais informações, consulte Criação de um banco de dados. Para criar uma tabela, uma entidade principal requer `CREATE_TABLE` permissão no banco de dados em que a tabela está sendo criada. Para obter mais informações, consulte Criar tabelas.

Permissões implícitas e explícitas

O Lake Formation fornece permissões implícitas, dependendo da persona e das ações que a persona realiza. Por exemplo, os administradores do data lake obtêm automaticamente permissões `DESCRIBE` para todos os recursos no catálogo de dados, permissões de localização de dados em todos os locais, permissões para criar bancos de dados e tabelas em todos os locais, bem como permissões `Grant` e `Revoke` em qualquer recurso. Os criadores de banco de dados obtêm automaticamente todas as permissões de banco de dados nos bancos de dados que eles criam, e os criadores de tabelas obtêm todas as permissões nas tabelas que eles criam. Para obter mais informações, consulte Permissões implícitas do Lake Formation.

Permissões concedidas

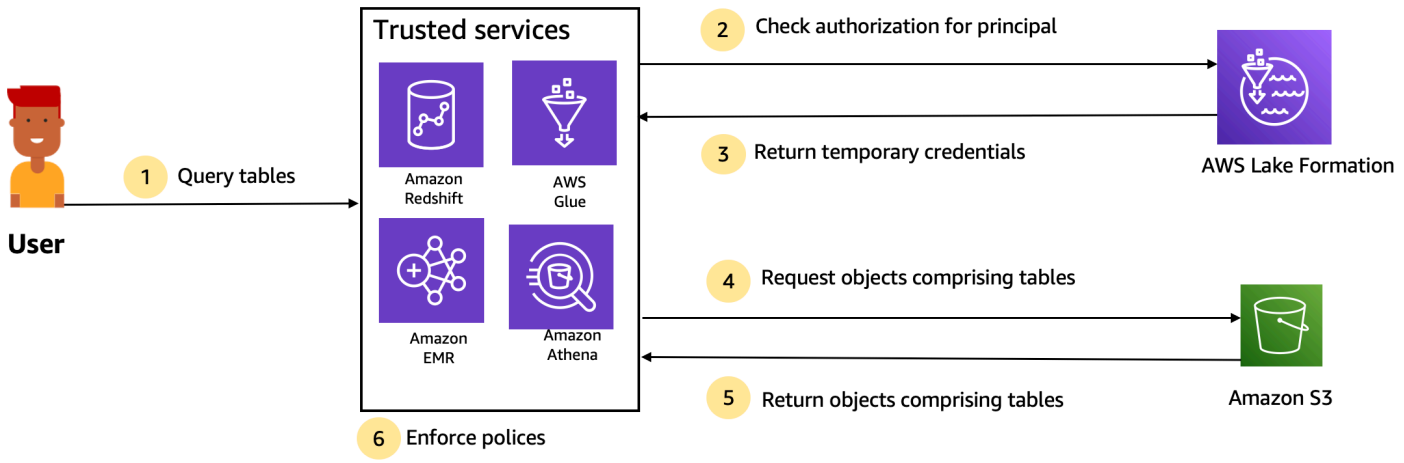
Os administradores do data lake têm a capacidade de delegar o gerenciamento de permissões a usuários não administrativos fornecendo permissões concedidas. Quando uma entidade principal recebe permissões concedidas sobre um recurso e um conjunto de permissões, essa entidade principal ganha a capacidade de conceder permissões a outras entidades principais nesse recurso.

Gerenciar o acesso às do armazenamento

O Lake Formation usa a funcionalidade de fornecimento de credenciais para conceder acesso temporário aos dados do Amazon S3. O fornecimento de credenciais, ou fornecimento de tokens, é um padrão comum que disponibiliza credenciais temporárias a usuários, serviços ou alguma outra entidade com o objetivo de conceder acesso de curto prazo a um recurso.

A Lake Formation aproveita esse padrão para fornecer acesso de curto prazo a serviços de AWS análise, como o Athena, para acessar dados em nome do responsável pela chamada. Ao conceder permissões, os usuários não precisam atualizar suas políticas de bucket do Amazon S3 ou políticas do IAM e não precisam de acesso direto ao Amazon S3.

O diagrama a seguir mostra como o Lake Formation fornece acesso temporário aos locais registrados:



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

1. Uma entidade principal (usuário) insere uma consulta ou solicitação de dados para uma tabela por meio de um serviço integrado confiável, como Athena, Amazon EMR, Redshift Spectrum ou o AWS Glue.
2. O serviço integrado verifica a autorização da Lake Formation para a tabela e as colunas solicitadas e determina a autorização. Se o usuário não estiver autorizado, o Lake Formation nega o acesso aos dados e a consulta falhará.
3. Depois que a autorização é bem-sucedida e a autorização de armazenamento é ativada para a tabela e o usuário, o serviço integrado recupera as credenciais temporárias do Lake Formation para acessar os dados.
4. O serviço integrado usa as credenciais temporárias do Lake Formation para solicitar objetos do Amazon S3.
5. O Amazon S3 fornece objetos do Amazon S3 para o serviço integrado. Os objetos do Amazon S3 contêm todos os dados da tabela.
6. O serviço integrado executa a aplicação necessária das políticas do Lake Formation, como filtragem em nível de coluna e nível de and/or célula em nível de linha. O serviço integrado processa as consultas e retorna os resultados ao usuário.

Habilite a aplicação de permissões em nível de armazenamento para tabelas do catálogo de dados

Por padrão, a imposição em nível de armazenamento não está habilitada para tabelas no catálogo de dados. Para permitir a fiscalização em nível de armazenamento, você deve registrar a localização dos seus dados de origem no Amazon S3 no Lake Formation e fornecer uma função do IAM.

Storage-level as permissões serão habilitadas para todas as tabelas com o mesmo caminho de localização da tabela ou prefixo da localização do Amazon S3.

Quando um serviço integrado solicita acesso ao local de dados em nome de um usuário, o serviço Lake Formation assume essa função e retorna as credenciais ao serviço solicitado com permissões reduzidas ao recurso para que o acesso aos dados possa ser feito. A função registrada do IAM deve ter todo o acesso necessário à localização do Amazon S3, incluindo AWS KMS as chaves.

Para obter mais informações, consulte Registrando uma localização do Amazon S3.

Compatível AWS serviços

AWS serviços analíticos, como Athena, Redshift Spectrum, Amazon AWS Glue EMR,, Amazon Quick, e se Amazon SageMaker AI integram ao AWS Lake Formation usando as operações da API de venda automática de credenciais do Lake Formation. Para ver uma lista completa dos AWS serviços que se integram ao Lake Formation e o nível de granularidade e os formatos de tabela compatíveis com eles, consulte. Trabalhando com outros AWS serviços

Cross-account compartilhamento de dados em Lake Formation

Com o Lake Formation, você pode compartilhar recursos do Catálogo de Dados (bancos de dados e tabelas) em uma AWS conta e entre contas em uma configuração simples usando o método de recurso nomeado ou LF-Tags. Você pode compartilhar um banco de dados inteiro ou selecionar tabelas de um banco de dados com qualquer entidade do IAM (funções e usuários do IAM) em uma conta, com outras AWS contas no nível da conta ou diretamente com entidades principais do IAM em outra conta.

Você também pode compartilhar tabelas do catálogo de dados com filtros de dados para restringir o acesso aos detalhes em nível de linha e de célula. Lake Formation usa AWS Resource Access Manager (AWS RAM) para facilitar a concessão de permissões entre contas. Quando um recurso é compartilhado entre duas contas, AWS RAM envia convites para a conta do destinatário. Quando um usuário aceita um convite de AWS RAM compartilhamento, AWS RAM fornece as permissões necessárias para que o Lake Formation tenha os recursos do Catálogo de Dados disponíveis, bem como habilite a imposição do nível de armazenamento. Para obter mais informações, consulte Cross-account compartilhamento de dados em Lake Formation.

Quando o administrador do data lake da conta do destinatário aceita o AWS RAM compartilhamento, os recursos compartilhados ficam disponíveis na conta do destinatário. O administrador do data lake concede mais permissões do Lake Formation no recurso compartilhado a outras entidades

principais do IAM na conta do destinatário, se o administrador tiver permissões `GRANTABLE` no recurso compartilhado.

No entanto, as entidades principais não podem consultar os recursos compartilhados usando o Athena ou o Redshift Spectrum sem um link de recurso. Um link de recurso é uma entidade no Catálogo de Dados e é semelhante a um Linux-Symlink conceito.

O administrador do data lake da conta do destinatário cria um link no recurso compartilhado. O administrador concede permissões `Describe` no link do recurso com as permissões necessárias no recurso compartilhado original para outros usuários. Um usuário na conta do destinatário pode então usar o link para consultar o recurso compartilhado usando o Athena e o Redshift Spectrum. Para obter mais informações sobre links de recursos, consulte Criação de links de recursos.

Componentes do Lake Formation

AWS Lake Formation depende da interação de vários componentes para criar e gerenciar seu data lake.

Lake Formation

Você usa o console do Lake Formation para definir e gerenciar seu data lake e conceder e revogar permissões do Lake Formation. Você pode usar esquemas no console para descobrir, limpar, transformar e ingerir dados. É possível habilitar ou desabilitar o acesso ao console para usuários individuais do Lake Formation.

API e interface de linha de comando do Lake Formation

Lake Formation fornece operações de API por meio de vários SDKs específicos de linguagem e do AWS Command Line Interface (AWS CLI). O Lake Formation API funciona em conjunto com a API do AWS Glue. A API Lake Formation se concentra principalmente no gerenciamento de permissões do Lake Formation, enquanto a API do AWS Glue fornece uma API do catálogo de dados e uma infraestrutura gerenciada para definir, programar e executar operações de ETL em seus dados.

Para obter informações sobre a API do AWS Glue, consulte o Guia do desenvolvedor do AWS Glue. Para obter informações sobre como usar a AWS CLI, consulte a Referência de comandos da AWS CLI.

Outros serviços da AWS

O Lake Formation usa os seguintes serviços:

- AWS Glue para orquestrar trabalhos e crawlers para transformar dados usando as transformações no AWS Glue.
- O IAM concederá políticas de permissões às entidades principais do Lake Formation. O modelo de permissão do Lake Formation aumenta o modelo de permissão do IAM para proteger seu data lake.

Terminologia do Lake Formation

A seguir estão alguns termos importantes que você encontrará neste guia.

Data lake

O data lake são seus dados persistentes que são armazenados no Amazon S3 e gerenciados pelo Lake Formation usando um catálogo de dados. Um data lake normalmente armazena o seguinte:

- Dados estruturados e não estruturados
- Dados brutos e dados transformados

Para que um caminho do Amazon S3 esteja dentro de um data lake, ele deve ser registrado com o Lake Formation.

Acesso aos dados

O Lake Formation fornece acesso seguro e granular aos dados por meio de um novo modelo de conceção/revogação de permissões que amplia as políticas do AWS Identity and Access Management (IAM).

Analistas e cientistas de dados podem usar o portfólio completo de serviços analíticos AWS e de machine learning, como o Amazon Athena, para acessar os dados. As políticas de segurança configuradas do Lake Formation ajudam a garantir que os usuários possam acessar somente os dados que estão autorizados a acessar.

Modo de acesso híbrido

O modo de acesso híbrido permite proteger e acessar os dados catalogados usando as permissões do Lake Formation e as permissões do IAM e do Amazon S3. O modo de acesso híbrido permite que os administradores de dados integrem as permissões do Lake Formation de forma seletiva e incremental, concentrando-se em um caso de uso do data lake por vez.

Blueprint

Um esquema é um modelo de gerenciamento de dados que permite a ingestão fácil de dados em um data lake. O Lake Formation fornece vários esquemas, um para cada tipo de fonte predefinido, como um banco de dados relacional ou logs do AWS CloudTrail. A partir de um esquema, você pode criar um fluxo de trabalho. Os fluxos de trabalho consistem em crawlers do AWS Glue, tarefas e acionadores que são gerados para orquestrar o carregamento e a atualização dos dados. Os esquemas usam a fonte de dados, o destino dos dados e o cronograma como entrada para configurar o fluxo de trabalho.

Fluxo de trabalho

Um fluxo de trabalho é um contêiner de um conjunto de tarefas relacionadas do AWS Glue, crawlers e gatilhos. Você cria o fluxo de trabalho no Lake Formation e ele é executado no serviço do AWS Glue. O Lake Formation pode rastrear o status de um fluxo de trabalho como uma entidade única.

Ao definir um fluxo de trabalho, você seleciona o esquema no qual ele se baseia. Em seguida, você pode executar fluxos de trabalho sob demanda ou de acordo com um cronograma.

Os fluxos de trabalho que você cria no Lake Formation são visíveis no console do AWS Glue como um gráfico acíclico direcionado (DAG). Ao usar o DAG, você pode acompanhar o andamento do fluxo de trabalho e solucionar o problema.

catálogo de dados

O catálogo de dados é seu armazenamento de metadados persistente. É um serviço gerenciado que permite armazenar, anotar e compartilhar metadados na Nuvem AWS da mesma maneira que você faria em uma metastore do Apache Hive. Ele fornece um repositório uniforme onde sistemas diferentes podem armazenar e encontrar metadados para rastrear dados em silos de dados e, em seguida, usar esses metadados para consultar e transformar os dados. O Lake Formation usa o catálogo de dados do AWS Glue para armazenar metadados sobre data lakes, fontes de dados, transformações e destinos.

Os metadados sobre fontes e destinos de dados estão na forma de bancos de dados e tabelas. As tabelas armazenam informações de esquemas, localização e muito mais. Bancos de dados são coleções de tabelas. O Lake Formation fornece uma hierarquia de permissões para controlar o acesso a bancos de dados e tabelas no catálogo de dados.

Cada conta da AWS tem um catálogo de dados por região da AWS.

Dados subjacentes

Os dados subjacentes se referem aos dados de origem ou aos dados dentro dos data lakes para os quais as tabelas do catálogo de dados apontam.

Entidade principal

Uma entidade principal é um usuário ou um perfil (IAM) AWS Identity and Access Management ou um usuário do Diretório Ativo.

Administrador do data lake

Um administrador de data lake é uma entidade principal que pode conceder a qualquer entidade principal (inclusive a si mesmo) qualquer permissão em qualquer recurso ou local de dados do catálogo de dados. Designe um administrador de data lake como o primeiro usuário do catálogo de dados. Esse usuário pode, então, conceder permissões mais granulares de recursos a outras entidades principais.

> **Note**
>
> Os usuários administrativos do IAM – usuários com a política `AdministratorAccess` gerenciada por AWS – não são automaticamente administradores de data lake. Por exemplo, eles não podem conceder permissões do Lake Formation em objetos do catálogo, a menos que tenham recebido permissão para fazer isso. No entanto, eles podem usar o console ou a API do Lake Formation para se designarem como administradores do data lake.

Para obter informações sobre os recursos de um administrador de data lake, consulte Permissões implícitas do Lake Formation. Para obter informações sobre como designar um usuário como administrador de data lake, consulte Crie um administrador de data lake.

AWS integrações de serviços com Lake Formation

Você pode usar o Lake Formation para gerenciar as permissões de acesso no nível do banco de dados, da tabela e da coluna sobre os dados armazenados no Amazon S3. Depois que seus dados forem registrados no Lake Formation, você poderá usar serviços AWS analíticos como Amazon Athena AWS Glue, Amazon Redshift Spectrum e Amazon EMR para consultar os dados. Os AWS serviços a seguir se integram AWS Lake Formation e honram as permissões do Lake Formation.

AWS Serviço	Os detalhes da integração
AWS Glue	Tópico de referência: Usando AWS Lake Formation com AWS Glue O AWS Glue e o Lake Formation compartilham o mesmo catálogo de dados. Para operações de console (como visualizar uma lista de tabelas) e todas as operações de API, os usuários do AWS Glue podem acessar somente os bancos de dados e tabelas nos quais têm permissões do Lake Formation.
Amazon Athena	Tópico de referência: Usando AWS Lake Formation com o Amazon Athena Usar o Lake Formation para permitir ou negar permissões para ler dados no Amazon S3. Quando os usuários do Amazon Athena selecionam o catálogo do AWS Glue no editor de consultas, eles podem consultar somente os bancos de dados, tabelas e colunas nos quais têm permissões do Lake Formation. Consultas usando manifestos não são aceitas. Atualmente, o Lake Formation não oferece suporte ao gerenciamento de permissões em operações de gravação como `VACUUM`, `MERGE`, `UPDATE` e `OPTIMIZE` em tabelas em formatos de tabela aberta. Além dos diretores que se autenticam com o Athena por meio do AWS Identity and Access Management (IAM), o Lake Formation oferece suporte aos usuários do Athena que se conectam por meio do driver JDBC ou ODBC e se autenticam por meio do SAML. Os provedores de SAML aceitos incluem o Okta e o Microsoft Active Directory Federation Service (AD FS).
Amazon Redshift Spectrum	Tópico de referência: Usando AWS Lake Formation com o Amazon Redshift Spectrum Quando os usuários do Amazon Redshift criam um esquema externo em um banco de dados no AWS Glue Data Catalog, eles podem consultar somente as tabelas e colunas desse esquema nas quais tenham permissões do Lake Formation.

AWS Serviço	Os detalhes da integração
Amazon Quick Enterprise Edition	Referência: Usando AWS Lake Formation com o Quick Quando um usuário do Amazon Quick Enterprise Edition consulta um conjunto de dados em um local do Amazon S3, o usuário deve ter a permissão Lake SELECT Formation nos dados.
Amazon EMR	Referência: Usando AWS Lake Formation com o Amazon EMR Você pode integrar as permissões do Lake Formation ao criar um cluster do Amazon EMR com uma função de runtime. Uma função de tempo de execução é uma função do IAM que você associa a trabalhos ou consultas do Amazon EMR e, em seguida, o Amazon EMR usa essa função para acessar recursos. AWS

O Lake Formation também trabalha com o AWS Key Management Service(AWS KMS) para permitir que você configure com mais facilidade esses serviços integrados para criptografar e descriptografar dados em locais do Amazon Simple Storage Service (Amazon S3).

Recursos adicionais do Lake Formation

Para obter mais informações sobre o AWS Lake Formation, recomendamos que continue aprendendo os conceitos apresentados neste guia usando os seguintes recursos:

Tópicos

- Blogs
- Palestras técnicas e webinars
- Arquitetura moderna
- Recursos de data mesh
- Guias de práticas recomendadas

Blogs

- AWS Lake Formation — Análise do ano de 2022

- Arquitetura de dados moderna e multirregional altamente resiliente
- Compartilhamento entre contas usando tags do LF para direcionar entidades principais do IAM
- Painel de inventário de permissões do Lake Formation
- Data mesh orientada por eventos

Palestras técnicas e webinars

- re:Invent 2020 — Data lakes: Crie, proteja e compartilhe com facilidade com AWS Lake Formation
- re:Invent 2022 — Construindo e operando um datalake no Amazon S3
- AWS Summit SF 2022 — Entendendo e alcançando uma arquitetura de dados moderna
- AWS Summit ATL 2022 — Data lakes modernos com o AWS Lake Formation, Amazon Redshift e AWS Glue
- AWS Summit ANZ 2022 — Data lakes, lake houses e data mesh: o quê, por quê e como?
- AWS Palestras técnicas online — Simplificando as permissões e a governança em seu data lake

Arquitetura moderna

- Padrões de arquitetura moderna

Recursos de data mesh

- Crie uma arquitetura de dados moderna e um padrão de data mesh em grande escala usando controle de acesso AWS Lake Formation baseado em tags
- Como o JPMorgan Chase construiu uma arquitetura de data mesh para gerar valor significativo e aprimorar sua plataforma de dados corporativos
- Crie uma data mesh no AWS

Guias de práticas recomendadas

- Guias de práticas recomendadas do AWS Lake Formation

Introdução ao Lake Formation

Recomendamos que você inicie por estas seções:

- AWS Lake Formation: como funciona: aprenda sobre a terminologia essencial e como os vários componentes interagem.
- Introdução ao Lake Formation: obtenha informações sobre os pré-requisitos e conclua tarefas importantes de configuração.
- AWS Lake Formation tutoriais— Siga os step-by-step tutoriais para aprender a usar o Lake Formation.
- Segurança em AWS Lake Formation: entenda como você pode ajudar a proteger o acesso a dados no Lake Formation.

Introdução ao Lake Formation

Se você não se inscreveu AWS ou não precisa de ajuda para começar, certifique-se de concluir as tarefas a seguir.

Tópicos

- Inicial completa AWS tarefas de configuração
- Configurar AWS Lake Formation
- Atualizando as permissões AWS Glue de dados para o modelo AWS Lake Formation
- AWS Lake Formation e endpoints VPC de interface ()AWS PrivateLink

Inicial completa AWS tarefas de configuração

Para usar o AWS Lake Formation, antes é necessário concluir as seguintes tarefas:

Tópicos

- Inscreva-se em um Conta da AWS

Inscreva-se em um Conta da AWS

Para começar AWS, você precisa de um Conta da AWS. Para obter informações sobre como criar um Conta da AWS, consulte Introdução a um Conta da AWS no Guia de AWS Gerenciamento de contas referência.

Configurar AWS Lake Formation

As seções a seguir fornecem informações sobre a configuração do Lake Formation pela primeira vez. Nem todos os tópicos desta seção são necessários para começar a usar o Lake Formation. Você pode usar as instruções para configurar o modelo de permissões do Lake Formation para gerenciar seus AWS Glue Data Catalog objetos e locais de dados existentes no Amazon Simple Storage Service (Amazon S3).

1. Crie um administrador de data lake
2. Alterar o modelo de permissão padrão ou usar o modo de acesso híbrido

3. the section called “Como configurar um local no Amazon S3 para o data lake”
4. the section called “Atribuir permissões aos usuários do Lake Formation”
5. the section called “Integrar o Centro de Identidade do IAM”
6. the section called “(Opcional) Configurações externas de filtragem de dados”
7. the section called “(Opcional) Conceder acesso à chave de criptografia do catálogo de dados”
8. (Opcional) Criar um perfil do IAM para fluxos de trabalho

Esta seção mostra como configurar os recursos do Lake Formation de duas maneiras diferentes:

- Usando um AWS CloudFormation modelo
- Como usar o console Lake Formation

Para configurar o Lake Formation usando o AWS console, acesseCrie um administrador de data lake.

Configure os recursos de formação de lagos usando CloudFormation modelo

> **Note**
>
> A CloudFormation pilha executa as etapas 1 a 6 acima, exceto as etapas 2 e 5. Execute Alterar o modelo de permissão padrão ou usar o modo de acesso híbrido e the section called “Integrar o Centro de Identidade do IAM” manualmente no console do Lake Formation.

1. Faça login no AWS CloudFormation console https://console.aws.amazon.com/cloudformation como administrador do IAM na região Leste dos EUA (Norte da Virgínia).
2. Selecione Iniciar Pilha.
3. Na página Criar pilha, selecione Avançar.
4. Digite um Nome de pilha.
5. Para DatalakeAdminName e DatalakeAdminPassword, digite seu nome de usuário e senha para o usuário administrador do data lake.
6. Para DatalakeUser1Name e DatalakeUser1Password, digite seu nome de usuário e senha para o usuário analista de data lake.

7. Para DataLakeBucketName, insira o nome do novo bucket que será criado.
8. Escolha Próximo.
9. Na próxima página, escolha `I acknowledge that CloudFormation might create IAM resources with custom names` e selecione Próximo.
10. Analise os detalhes na página final e selecione Eu reconheço que isso AWS CloudFormation pode criar recursos do IAM.
11. Escolha Criar.

    A criação da pilha pode levar até dois minutos.

Limpar os recursos

Se você quiser limpar os recursos da CloudFormation pilha:

1. De-register o bucket do Amazon S3 que sua pilha criou e registrou como um local de data lake.
2. Exclua a CloudFormation pilha. Isso excluirá todos os recursos criados pela pilha.

Crie um administrador de data lake

Inicialmente, os administradores de data lake são os únicos usuários ou funções AWS Identity and Access Management (IAM) que podem conceder permissões do Lake Formation em locais de dados e recursos do catálogo de dados a qualquer diretor (inclusive a si mesmo). Para obter mais informações sobre os recursos do administrador de data lake, consulte Permissões implícitas do Lake Formation. Por padrão, o Lake Formation permite criar até 30 administradores de data lake.

Você pode criar um administrador de data lake usando o console do Lake Formation ou a operação `PutDataLakeSettings` da API do Lake Formation.

As permissões a seguir são necessárias para criar um administrador de data lake. O usuário `Administrator` tem essas permissões implicitamente.

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

Se você conceder a política `AWSLakeFormationDataAdmin` a um usuário, esse usuário não poderá criar usuários administradores adicionais do Lake Formation.

Como criar um administrador de data lake (console)

1. Se o usuário que será administrador do data lake ainda não existir, use o console do IAM para criá-lo. Caso contrário, selecione um usuário existente que será o administrador do data lake.

   > **Note**
   >
   > Recomendamos que você não selecione um usuário administrativo do IAM (usuário com a política `AdministratorAccess` AWS gerenciada) para ser o administrador do data lake.

   Anexe as seguintes políticas AWS gerenciadas ao usuário:

Políticas	Obrigatório?	Observações
`AWSLakeFormationDataAdmin`	Obrigatório	Permissões básicas do administrador do data lake. Essa política AWS gerenciada contém uma negação explícita da operação da API Lake Formation, `PutDataLakeSetting` que impede que os usuários criem novos administradores de data lake.
`AWSGlueConsoleFullAccess` , `CloudWatchLogsReadOnlyAccess`	Opcional	Anexe essas políticas se o administrador do data lake estiver solucionando problemas de fluxos de trabalho criados a partir dos esquemas do Lake Formation. Essas políticas permitem que o administrador do data lake visualize as informações de solução de problemas no console do AWS Glue e no console do Amazon CloudWatch Logs . Para obter informações sobre fluxos de trabalho, consulte the section called “Importação de dados usando fluxos de trabalho”.

Políticas	Obrigatório?	Observações
`AWSLakeFormationCrossAccountManager`	Opcional	Anexe esta política para permitir que o administrador do data lake conceda e revogue permissões entre contas nos recursos do catálogo de dados. Para obter mais informações, consulte Cross-account compartilhamento de dados em Lake Formation.
`AmazonAthenaFullAccess`	Opcional	Anexe essa política se o administrador do data lake estiver executando consultas em Amazon Athena.

2. Anexe a seguinte política em linha, que concede ao administrador do data lake permissão para criar o perfil vinculado ao serviço do Lake Formation. O nome sugerido para a política é `LakeFormationSLR`.

   O perfil vinculado ao serviço permite que o administrador do data lake registre com mais facilidade o local do Amazon S3 no Lake Formation. Para obter mais informações sobre o perfil vinculado ao serviço Lake Formation, consulte the section called “Uso de perfis vinculados ao serviço”.

   > **Important**
   >
   > Em todas as políticas a seguir, *<account-id>* substitua por um número de AWS conta válido.

```
{
    "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. (Opcional) Anexe a seguinte política em linha ao usuário `PassRole`. Essa política permite que o administrador do data lake crie e execute fluxos de trabalho. A permissão `iam:PassRole` permite que o fluxo de trabalho assuma o perfil `LakeFormationWorkflowRole` para criar crawlers e trabalhos, e anexe o perfil aos crawlers e trabalhos criados. O nome sugerido para a política é `UserPassRole`.

> **Important**
>
> *<account-id>*Substitua por um número de AWS conta válido.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "PassRolePermissions",
            "Effect": "Allow",
            "Action": [
                "iam:PassRole"
            ],
            "Resource": [
                "arn:aws:iam::111122223333:role/LakeFormationWorkflowRole"
            ]
        }
    ]
```

```
}
```

4. (Opcional) Anexe essa política adicional em linha se sua conta estiver concedendo ou recebendo permissões entre contas do Lake Formation. Essa política permite que o administrador do data lake visualize e aceite AWS Resource Access Manager (AWS RAM) convites de compartilhamento de recursos. Além disso, para administradores de data lake na conta AWS Organizations de gerenciamento, a política inclui uma permissão para permitir concessões entre contas para organizações. Para obter mais informações, consulte Cross-account compartilhamento de dados em Lake Formation.

   O nome sugerido para a política é `RAMAccess`.

   JSON

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

5. Abra o AWS Lake Formation console em https://console.aws.amazon.com/lakeformation/ e faça login como usuário com a política `AdministratorAccess` AWS gerenciada.
6. Se a janela de Boas-vindas ao Lake Formation for exibida, escolha o usuário do IAM que você criou ou selecionou na Etapa 1 e, em seguida, escolha Começa.
7. Se você não vir a janela de Boas-vindas ao Lake Formation, execute as etapas a seguir para configurar um administrador do Lake Formation.

   a. No painel de navegação, em Administração, selecione Perfis e tarefas administrativas. Na seção Administradores do Data Lake da página do console, selecione Adicionar.

b. Na caixa de diálogo Adicionar administradores, em Tipo de Acesso, selecione Administrador do Data Lake.

c. Para Usuários e perfis do IAM, selecione o usuário do IAM que você criou ou selecionou na Etapa 1 e, em seguida, selecione Salvar.

Alterar o modelo de permissão padrão ou usar o modo de acesso híbrido

O Lake Formation começa com as configurações “Usar somente controle de acesso IAM” habilitadas para compatibilidade com o AWS Glue Data Catalog comportamento existente. Essas configurações permitem que você gerencie o acesso aos seus dados no data lake e seus metadados por meio de políticas do IAM e políticas de bucket do Amazon S3.

Para facilitar a transição das permissões do data lake de um modelo do IAM e do Amazon S3 para as permissões do Lake Formation, recomendamos que você use o modo de acesso híbrido para o catálogo de dados. Com o modo de acesso híbrido, você tem um caminho incremental em que pode habilitar as permissões do Lake Formation para um conjunto específico de usuários sem interromper outros usuários ou workloads existentes.

Para obter mais informações, consulte Modo de acesso híbrido.

Desative as configurações padrão para mover todos os usuários existentes de uma tabela para o Lake Formation em uma única etapa.

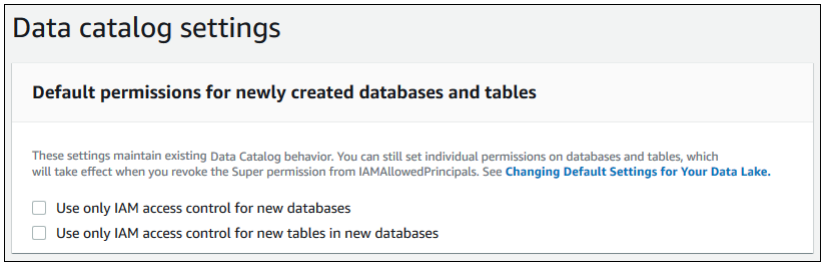
> **Important**
>
> Se você tiver bancos de dados AWS Glue Data Catalog e tabelas existentes, não siga as instruções descritas nesta seção. Em vez disso, siga as instruções em the section called “Atualizar as permissões de dados AWS Glue para o modelo do Lake Formation”.

> **Warning**
>
> Se você tiver uma automação que cria bancos de dados e tabelas no catálogo de dados, as etapas a seguir podem fazer com que as tarefas de automação e extração, transformação e carregamento (ETL) posteriores falhem. Prossiga somente depois de modificar seus processos existentes ou conceder permissões explícitas do Lake Formation às entidades

principais necessárias. Para obter informações sobre as permissões do Lake Formation, consulte the section called “Referência de permissões do Lake Formation”.

Para alterar as configurações padrão do catálogo de dados

1. Continue no console do Lake Formation em https://console.aws.amazon.com/lakeformation/. Certifique-se de estar conectado como usuário com a política `AdministratorAccess` AWS gerenciada.
2. Modificar as configurações do catálogo de dados:
   a. No painel de navegação, em Administração, selecione Configurações do catálogo de dados.
   b. Desmarque as duas caixas de seleção e selecione Salvar.



3. Revogar a permissão `IAMAllowedPrincipals` para criadores de banco de dados.
   a. No painel de navegação, em Administração, selecione Perfis e tarefas administrativas.
   b. Na página do console de Perfis e tarefas administrativas, na seção Criadores de banco de dados, selecione o grupo `IAMAllowedPrincipals`, e selecione Revogar.

      A caixa de diálogo Revogar permissões é exibida, mostrando que `IAMAllowedPrincipals` tem a permissão para Criar banco de dados.
   c. Selecione Revogar.

Atribuir permissões aos usuários do Lake Formation

Crie um usuário para ter acesso ao data lake em AWS Lake Formation. Esse usuário tem as permissões de privilégio mínimo para consultar o data lake.

Para obter mais informações sobre a criação de usuários ou grupos, consulte Identidades do IAM no Guia do usuário do IAM.

Como anexar permissões a um usuário não administrador para acessar os dados do Lake Formation

1. Abra o console do IAM em https://console.aws.amazon.com/iam e faça login como usuário com a política `AdministratorAccess` AWS gerenciada.
2. Selecione Usuários ou Grupos de usuários.
3. Na lista, selecione o nome do usuário ou do grupo ao qual deseja incorporar uma política.

   Selecione Permissões.
4. Selecione Adicionar permissões e selecione Anexar políticas diretamente. Digite `Athena` no campo de texto Políticas de filtro. Na lista de resultados, marque a caixa de seleção `AmazonAthenaFullAccess`.
5. Selecione o botão Criar política. Na página Criar política, selecione a guia JSON. Copie e cole o código a seguir no editor de políticas.

   JSON

```
{
    "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. Escolha Avançar até ver a página Revisar política. Digite um nome para a política, por exemplo, `DatalakeUserBasic`. Selecione Criar política e feche a guia Políticas ou a janela do navegador.

Como configurar um local no Amazon S3 para o data lake

Para usar o Lake Formation para gerenciar e proteger os dados em seu data lake, você deve primeiro registrar um local no Amazon S3. Quando você registra um local, esse caminho do Amazon S3 e todas as pastas sob esse caminho são registradas, o que permite que o Lake Formation aplique permissões de nível de armazenamento. Quando o usuário solicita dados de um mecanismo integrado como o Amazon Athena, o Lake Formation fornece acesso aos dados em vez de usar as permissões do usuário.

Ao registrar um local, você especifica uma função do IAM que concede read/write permissões nesse local. A Lake Formation assume essa função ao fornecer credenciais temporárias para AWS serviços integrados que solicitam acesso aos dados no local registrado do Amazon S3. É possível especificar o perfil vinculado ao serviço (SLR) do Lake Formation ou criar seu próprio perfil.

Use um perfil personalizado nas seguintes situações:

- Você planeja publicar métricas no Amazon CloudWatch Logs. A função definida pelo usuário deve incluir uma política para adicionar registros aos CloudWatch registros e publicar métricas, além das permissões de SLR. Para ver um exemplo de política em linha que concede CloudWatch as permissões necessárias, consulteRequisitos para funções usadas para registrar locais.
- O local do Amazon S3 existe em uma conta diferente. Para obter detalhes, consulte the section called “Registrando uma localização do Amazon S3 em outra AWS account”.
- O local do Amazon S3 contém dados criptografados com uma Chave gerenciada pela AWS. Para obter mais detalhes, consulte Registrando uma localização criptografada do Amazon S3 e Registrando uma localização criptografada do Amazon S3 em AWS contas.
- Você planeja acessar o local do Amazon S3 usando o Amazon EMR. Para obter mais informações sobre os requisitos de perfil, consulte Perfis do IAM para o Lake Formation no Guia de Gerenciamento do Amazon EMR.

O perfil que você escolher deve ter as permissões necessárias, conforme descrito em Requisitos para funções usadas para registrar locais. Para obter instruções sobre como registrar um local no Amazon S3, consulte Adicionar uma localização do Amazon S3 ao seu data lake.

(Opcional) Configurações externas de filtragem de dados

Se você pretende analisar e processar dados em seu data lake usando mecanismos de consulta de terceiros, você deve optar por permitir que mecanismos externos acessem dados gerenciados pelo Lake Formation. Se você não optar por participar, mecanismos externos não poderão acessar dados em locais do Amazon S3 registrados no Lake Formation.

O Lake Formation oferece suporte a permissões em nível de coluna para restringir o acesso a colunas específicas em uma tabela. Serviços analíticos integrados Amazon Athena, como Amazon Redshift Spectrum e Amazon EMR, recuperam metadados de tabelas não filtrados do. AWS Glue Data Catalog A filtragem real das colunas nas respostas da consulta é de responsabilidade do serviço integrado. É responsabilidade dos administradores terceirizados lidar adequadamente com as permissões para evitar o acesso não autorizado aos dados.

Para optar por permitir que mecanismos de terceiros acessem e filtrem dados (console)

1. Continue no console do Lake Formation em https://console.aws.amazon.com/lakeformation/. Certifique-se de estar conectado como entidade principal com a permissão do IAM na operação da API `PutDataLakeSettings` do Lake Formation. O usuário administrador do IAM que você criou em Inscreva-se em um Conta da AWS tem essa permissão.
2. No painel de navegação, em Administração, selecione Configurações de integração de aplicativos.
3. Na página Configurações de integração de aplicativos, faça o seguinte:

    a. Marque a caixa Permitir que mecanismos externos filtrem dados em locais do Amazon S3 registrados no Lake Formation.

    b. Digite os Valores da tag de sessão definidos para mecanismos de terceiros.

    c. Em IDs de conta da AWS , digite os IDs de conta de onde mecanismos de terceiros têm permissão para acessar locais registrados com o Lake Formation. Pressione Enter após cada ID da conta.

    d. Selecione Salvar.

Para permitir que mecanismos externos acessem dados sem a validação da tag de sessão, consulte Integração de aplicativos para acesso total à tabela

(Opcional) Conceder acesso à chave de criptografia do catálogo de dados

Se o AWS Glue Data Catalog estiver criptografado, conceda permissões AWS Identity and Access Management (IAM) na AWS KMS chave a qualquer diretor que precise conceder permissões de Lake Formation em bancos de dados e tabelas do Catálogo de Dados.

Para obter mais informações, consulte o Guia do desenvolvedor do AWS Key Management Service .

(Opcional) Criar um perfil do IAM para fluxos de trabalho

Com AWS Lake Formation, você pode importar seus dados usando fluxos de trabalho executados por AWS Glue rastreadores. Um fluxo de trabalho define a fonte de dados e o cronograma para importar dados para o seu data lake. Você pode definir facilmente fluxos de trabalho usando os esquemas ou modelos fornecidos pelo Lake Formation.

Ao criar um fluxo de trabalho, você deve atribuir a ele uma função AWS Identity and Access Management (IAM) que conceda à Lake Formation as permissões necessárias para ingerir os dados.

O procedimento a seguir requer familiaridade com o IAM.

Para criar um perfil do IAM para fluxos de trabalho

1. Abra o console do IAM em https://console.aws.amazon.com/iam e faça login como usuário com a política `AdministratorAccess` AWS gerenciada.
2. No painel de navegação, selecione Perfis e depois Criar perfil.
3. Na página Criar perfil, selecione Serviço da AWS e, em seguida, selecione Glue. Escolha Próximo.
4. Na página Adicionar permissões, pesquise a política AWSGlueServiceRole gerenciada e marque a caixa de seleção ao lado do nome da política na lista. Em seguida, conclua o assistente de Criação de perfil, nomeando o perfil`LFWorkflowRole`. Para finalizar, selecione Criar perfil.
5. Na página Perfis, procure seu perfil por `LFWorkflowRole` e selecione o nome.
6. Na página Resumo do perfil, na guia Permissões, selecione Adicionar política em linha. Na tela Criar política, navegue até a guia JSON e adicione a seguinte política em linha. O nome sugerido para a política é `LakeFormationWorkflow`.

> **Important**
>
> Na política a seguir, *<account-id>* substitua por um Conta da AWS número válido.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                 "lakeformation:GetDataAccess",
                 "lakeformation:GrantPermissions"
             ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": ["iam:PassRole"],
            "Resource": [
                "arn:aws:iam::111122223333:role/LakeFormationWorkflowRole"
            ]
        }
    ]
}
```

Seguem breves descrições das permissões nesta política:

- `lakeformation:GetDataAccess` permite que trabalhos criados pelo fluxo de trabalho sejam gravados no local de destino.
- `lakeformation:GrantPermissions` permite que o fluxo de trabalho conceda a `SELECT` permissão nas tabelas de destino.
- `iam:PassRole` permite que o serviço assuma o perfil de `LakeFormationWorkflowRole` para criar crawlers e trabalhos (instâncias de fluxos de trabalho) e anexe o perfil aos crawlers e trabalhos criados.

7. Verifique se o perfil `LakeFormationWorkflowRole` tem duas políticas anexadas.
8. Se você estiver ingerindo dados fora do local do data lake, adicione uma política em linha que conceda permissões para ler os dados de origem.

Atualizando as permissões AWS Glue de dados para o modelo AWS Lake Formation

AWS Lake Formation as permissões permitem um controle de acesso refinado para dados em seu data lake. Você pode usar o modelo de permissões do Lake Formation para gerenciar seus AWS Glue Data Catalog objetos e locais de dados existentes no Amazon Simple Storage Service (Amazon S3).

O modelo de permissões do Lake Formation usa permissões granulares AWS Identity and Access Management (IAM) para acesso ao serviço de API. O Lake Formation usa a funcionalidade Filtragem de dados e segurança por célula no Lake Formation para restringir o acesso à tabela em nível de coluna, linha e célula para os usuários e as respectivas aplicações. Em comparação, o modelo do AWS Glue concede acesso aos dados por meio de políticas do IAM baseadas em identidade e recursos.

Para fazer a troca, siga as etapas deste guia.

Para obter mais informações, consulte Visão geral das permissões do Lake Formation.

Sobre permissões padrão

Para manter a compatibilidade com versões anterioresAWS Glue, por padrão, AWS Lake Formation concede a `Super` permissão ao `IAMAllowedPrincipals` grupo em todos os recursos existentes do Catálogo de AWS Glue Dados e concede a `Super` permissão em novos recursos do Catálogo de Dados se as configurações de controle de acesso Use only IAM estiverem ativadas. Isso efetivamente faz com que o acesso aos recursos do catálogo de dados e aos locais do Amazon S3 seja controlado exclusivamente pelas políticas do AWS Identity and Access Management (IAM). O grupo `IAMAllowedPrincipals` inclui todos os usuários e funções do IAM que têm permissão para acessar seus objetos do catálogo de dados por meio de suas políticas do IAM. A permissão `Super` possibilita que uma entidade principal execute todas as operações suportadas do Lake Formation no banco de dados ou na tabela em que ela foi concedida.

Você pode começar a usar o Lake Formation para gerenciar o acesso aos seus dados registrando os locais dos recursos existentes do catálogo de dados no Lake Formation ou usando o modo de

acesso híbrido. Ao registrar a localização do Amazon S3 no modo de acesso híbrido, você pode habilitar as permissões do Lake Formation optando por entidades principais para bancos de dados e tabelas nesse local.

Para facilitar a transição das permissões do data lake de um modelo do IAM e do Amazon S3 para as permissões do Lake Formation, recomendamos que você use o modo de acesso híbrido para o catálogo de dados. Com o modo de acesso híbrido, você tem um caminho incremental em que pode habilitar as permissões do Lake Formation para um conjunto específico de usuários sem interromper outros usuários ou workloads existentes.

Para obter mais informações, consulte Modo de acesso híbrido.

Desative as configurações padrão do catálogo de dados para transferir todos os usuários existentes de uma tabela para o Lake Formation em uma única etapa.

Para começar a usar as permissões do Lake Formation com bancos de dados e tabelas do catálogo de dados do AWS Glue existentes, você deve fazer o seguinte:

1. Determine as permissões do IAM existentes dos seus usuários para cada banco de dados e tabela.
2. Replique essas permissões no Lake Formation.
3. Para cada local do Amazon S3 que contém dados:
   a. Revogue a permissão `Super` do grupo `IAMAllowedPrincipals` em cada recurso do catálogo de dados que faça referência a esse local.
   b. Registre o local com o Lake Formation.
4. Limpe as políticas do IAM de controle de acesso fino existentes.

> **Important**
>
> Para adicionar novos usuários durante o processo de transição do seu catálogo de dados, você deve configurar permissões granulares do AWS Glue no IAM como antes. Você também deve replicar essas permissões no Lake Formation conforme descrito nesta seção. Se os novos usuários tiverem as políticas gerais do IAM descritas neste guia, eles poderão listar quaisquer bancos de dados ou tabelas que tenham a permissão `Super` concedida para `IAMAllowedPrincipals`. Eles também podem visualizar os metadados desses recursos.

Siga as etapas desta seção para atualizar para o modelo de permissões do Lake Formation.

Tópicos

- Etapa 1: listar as permissões existentes dos usuários e das funções
- Etapa 2: configurar permissões equivalentes do Lake Formation
- Etapa 3: conceder aos usuários permissões do IAM para usar o Lake Formation
- Etapa 4: mude seus armazenamentos de dados para o modelo de permissões do Lake Formation
- Etapa 5: proteja os novos recursos do catálogo de dados
- Etapa 6: fornecer aos usuários uma nova política do IAM para acesso futuro ao data lake
- Etapa 7: limpar políticas do IAM existentes

Etapa 1: listar as permissões existentes dos usuários e das funções

Para começar a usar AWS Lake Formation permissões com seus AWS Glue bancos de dados e tabelas existentes, você deve primeiro determinar as permissões existentes dos seus usuários.

> **Important**
>
> Antes de começar, realize as seguintes tarefas em Introdução.

Tópicos

- Usar a operação da API
- Usando o Console de gerenciamento da AWS
- Usando AWS CloudTrail

Usar a operação da API

Use a operação da ListPoliciesGrantingServiceAccessAPI AWS Identity and Access Management (IAM) para determinar as políticas do IAM anexadas a cada principal (usuário ou função). A partir das políticas retornadas nos resultados, você pode determinar as permissões do IAM que são concedidas à entidade principal. Você deve invocar a API para cada entidade principal separadamente.

Example

O AWS CLI exemplo a seguir retorna as políticas anexadas ao usuário`glue_user1`.

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

O comando retorna resultados semelhantes ao seguinte.

```
{
    "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
}
```

Usando o Console de gerenciamento da AWS

Você também pode ver essas informações no console AWS Identity and Access Management (IAM), na guia Access Advisor na página de resumo do usuário ou da função:

1. Abra o console do IAM em https://console.aws.amazon.com/iam/.
2. No painel de navegação, selecione Usuários ou Funções.
3. Escolha um nome na lista para abrir a página Resumo e escolha a guia Supervisor de Acesso.
4. Inspecione cada uma das políticas para determinar a combinação de bancos de dados, tabelas e ações para as quais cada usuário tem permissões.

Lembre-se de inspecionar as funções, além dos usuários, durante esse processo, pois seus trabalhos de processamento de dados podem estar assumindo funções para acessar os dados.

Usando AWS CloudTrail

Outra forma de determinar suas permissões existentes é procurar chamadas de AWS Glue API em AWS CloudTrail que o `additionaleventdata` campo dos registros contenha uma `insufficientLakeFormationPermissions` entrada. Essa entrada lista o banco de dados e a tabela nos quais o usuário precisa das permissões do Lake Formation para realizar a mesma ação.

Esses são logs de acesso a dados, portanto, não é garantido que produzam uma lista abrangente de usuários e suas permissões. Recomendamos escolher um intervalo de tempo amplo para capturar a maioria dos padrões de acesso aos dados de seus usuários, por exemplo, várias semanas ou meses.

Para obter mais informações, consulte Visualização de CloudTrail eventos com histórico de eventos no Guia AWS CloudTrail do usuário.

Em seguida, você pode configurar as permissões do Lake Formation de acordo com as permissões do AWS Glue. Consulte Etapa 2: configurar permissões equivalentes do Lake Formation.

Etapa 2: configurar permissões equivalentes do Lake Formation

Usando as informações coletadas emEtapa 1: listar as permissões existentes dos usuários e das funções, conceda AWS Lake Formation permissões que correspondam às AWS Glue permissões. Use qualquer um dos métodos a seguir para realizar as concessões:

- Você pode usar o console do Lake Formation ou a AWS CLI.

  Consulte the section called “Conceder permissões de dados”.
- Use as operações da API `GrantPermissions` ou `BatchGrantPermissions`.

  Consulte APIs de permissões.

Para obter mais informações, consulte Visão geral das permissões do Lake Formation.

Depois de configurar as permissões do Lake Formation, siga para a Etapa 3: conceder aos usuários permissões do IAM para usar o Lake Formation.

Etapa 3: conceder aos usuários permissões do IAM para usar o Lake Formation

Para usar o modelo de AWS Lake Formation permissões, os diretores devem ter permissões AWS Identity and Access Management (IAM) no Lake Formation APIs.

Crie a política a seguir no IAM e a anexe a todos os usuários que precisam acessar seu data lake. Atribua o nome `LakeFormationDataAccess` à política.

JSON

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

Em seguida, atualize para as permissões do Lake Formation, um local de dados por vez. Consulte Etapa 4: mude seus armazenamentos de dados para o modelo de permissões do Lake Formation.

Etapa 4: mude seus armazenamentos de dados para o modelo de permissões do Lake Formation

Atualize para as permissões do Lake Formation, um local de dados por vez. Para isso, repita essa seção inteira até ter registrado todos os caminhos do Amazon Simple Storage Service (Amazon S3) referenciados pelo catálogo de dados.

Tópicos

- Verifique permissões do Lake Formation

- Proteja os recursos existentes do catálogo de dados
- Ative as permissões do Lake Formation para sua localização no Amazon S3

Verifique permissões do Lake Formation

Antes de registrar um local, execute uma etapa de verificação para garantir que as entidades principais corretas tenham as permissões necessárias para o Lake Formation e que nenhuma permissão para o Lake Formation seja concedida às entidades principais que não deveriam tê-las. Ao usar a operação da API `GetEffectivePermissionsForPath` do Lake Formation, identifique os recursos do catálogo de dados que fazem referência à localização do Amazon S3, junto com as entidades principais que têm permissões sobre esses recursos.

O AWS CLI exemplo a seguir retorna os bancos de dados e tabelas do catálogo de dados que fazem referência ao bucket do Amazon S3. `products`

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

Observe a opção `profile`. Recomendamos que você execute o comando como administrador do data lake.

A seguir está um trecho dos resultados retornados.

```
{
        "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**
>
> Se o catálogo de dados do AWS Glue estiver criptografado, `GetEffectivePermissionsForPath` retornará somente bancos de dados e tabelas que foram criados ou modificados após a disponibilidade geral do Lake Formation.

Proteja os recursos existentes do catálogo de dados

Em seguida, revogue a permissão `Super` de cada tabela e banco de dados `IAMAllowedPrincipals` que você identificou para o local.

> **Warning**
>
> Se você tiver uma automação que cria bancos de dados e tabelas no catálogo de dados, as etapas a seguir podem fazer com que as tarefas de automação e extração, transformação e carregamento (ETL) posteriores falhem. Prossiga somente depois de modificar seus processos existentes ou conceder permissões explícitas do Lake Formation às entidades principais necessárias. Para obter informações sobre as permissões do Lake Formation, consulte the section called “Referência de permissões do Lake Formation”.

Para revogar **`Super`** de **`IAMAllowedPrincipals`** em uma tabela

1. Abra o AWS Lake Formation console em https://console.aws.amazon.com/lakeformation/. Faça login como administrador de data lake.
2. No painel de navegação, selecione Tabelas.
3. Na página Tabelas, selecione o botão de opção ao lado da tabela desejada.
4. No menu Ações, selecione Revogar.
5. Na caixa de diálogo Revogar permissões, na lista de usuários e funções do IAM, role para baixo até o título Grupo e escolha IAMAllowedPrincipals.
6. Em Permissões da tabela, verifique se a opção Super está selecionada e escolha Revogar.

Para revogar **`Super`** de **`IAMAllowedPrincipals`** em um banco de dados

1. Abra o AWS Lake Formation console em https://console.aws.amazon.com/lakeformation/. Faça login como administrador de data lake.
2. No painel de navegação, escolha Bancos de dados.
3. Na página Banco de dados, selecione o botão de opção ao lado do banco de dados desejado.
4. No menu Ações, escolha Editar.
5. Na página Editar banco de dados, desmarque Usar somente o controle de acesso do IAM para novas tabelas nesse banco de dados e escolha Salvar.
6. De volta à página Banco de dados, verifique se o banco de dados ainda está selecionado e, no menu Ações, escolha Revogar.
7. Na caixa de diálogo Revogar permissões, na lista de usuários e funções do IAM, role para baixo até o título Grupo e escolha IAMAllowedPrincipals.
8. Em Permissões de banco de dados, verifique se a opção Super está selecionada e escolha Revogar.

Ative as permissões do Lake Formation para sua localização no Amazon S3

Em seguida, registre o local do Amazon S3 com o Lake Formation. Para isso, você pode usar o processo descrito em Adicionar uma localização do Amazon S3 ao seu data lake. Ou utilize a operação da API `RegisterResource`, conforme descrito em APIs de fornecimento de credenciais.

> **Note**
>
> Se o local pai estiver registrado, você não precisa registrar locais filho.

Depois de concluir essas etapas e testar se seus usuários podem acessar seus dados, você atualizou com sucesso para as permissões do Lake Formation. Continue na próxima etapa, Etapa 5: proteja os novos recursos do catálogo de dados.

Etapa 5: proteja os novos recursos do catálogo de dados

Em seguida, assegure todos os novos recursos do catálogo de dados alterando as configurações padrão do catálogo de dados. Desative as opções de usar somente o controle de acesso AWS Identity and Access Management (IAM) para novos bancos de dados e tabelas.

> **Warning**
>
> Se você tiver uma automação que cria bancos de dados e tabelas no catálogo de dados, as etapas a seguir podem fazer com que as tarefas de automação e extração, transformação e carregamento (ETL) posteriores falhem. Prossiga somente depois de modificar seus processos existentes ou conceder permissões explícitas do Lake Formation às entidades principais necessárias. Para obter informações sobre as permissões do Lake Formation, consulte the section called “Referência de permissões do Lake Formation”.

Para alterar as configurações padrão do catálogo de dados

1. Abra o AWS Lake Formation console em https://console.aws.amazon.com/lakeformation/. Faça login como usuário administrativo do IAM (o usuário `Administrator` ou outro usuário com a política `AdministratorAccess` AWS gerenciada).
2. No painel de navegação, selecione Configurações.
3. Na página Configurações do catálogo de dados, desmarque as duas caixas de seleção e escolha Salvar.

A próxima etapa é conceder aos usuários acesso a bancos de dados ou tabelas adicionais no futuro. Consulte Etapa 6: fornecer aos usuários uma nova política do IAM para acesso futuro ao data lake.

Etapa 6: fornecer aos usuários uma nova política do IAM para acesso futuro ao data lake

Para conceder aos seus usuários acesso a bancos de dados ou tabelas adicionais do Data Catalog no futuro, você deve fornecer a eles a política embutida de granularidade geral AWS Identity and Access Management (IAM) que segue. Atribua o nome `GlueFullReadAccess` à política.

> **Important**
>
> Se você anexar essa política a um usuário antes da revogação de `Super` de `IAMAllowedPrincipals` em cada banco de dados e tabela em seu catálogo de dados, esse usuário poderá visualizar todos os metadados de qualquer recurso ao qual `Super` tenha sido concedido a `IAMAllowedPrincipals`.

JSON

```
{
    "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**
>
> As políticas em linha designadas nesta etapa e nas etapas anteriores contêm permissões mínimas do IAM. Para políticas sugeridas para administradores de data lake, analistas de dados e outras pessoas, consulte the section called “Referência de personas e permissões do IAM do Lake Formation”.

Em seguida, prossiga para a Etapa 7: limpar políticas do IAM existentes.

Etapa 7: limpar políticas do IAM existentes

Depois de configurar as AWS Lake Formation permissões e criar e anexar as políticas gerais de controle de acesso AWS Identity and Access Management (IAM), conclua a seguinte etapa final:

- Remova dos usuários, grupos e funções que as antigas políticas de controle de acesso refinado do IAM que você replicou no Lake Formation.

Ao fazer isso, você garante que essas entidades principais não tenham mais acesso direto aos dados no Amazon Simple Storage Service (Amazon S3). Em seguida, você pode gerenciar o acesso dessas entidades principais ao data lake inteiramente por meio do Lake Formation.

AWS Lake Formation e endpoints VPC de interface ()AWS PrivateLink

O Amazon VPC é um AWS serviço que você pode usar para lançar AWS recursos em uma rede virtual que você define. Com a VPC, você tem controle sobre as configurações de rede, como o intervalo de endereços IP, sub-redes, tabelas de rotas e gateways de rede.

Se você usa a Amazon Virtual Private Cloud (Amazon VPC) para hospedar seus AWS recursos, você pode estabelecer uma conexão privada entre sua VPC e a Lake Formation. Você usa essa conexão para que o Lake Formation possa se comunicar com os recursos em sua VPC sem passar pela Internet pública.

Você pode estabelecer uma conexão privada entre sua VPC e criar uma AWS Lake Formation interface VPC endpoint. Os endpoints de interface são alimentados por AWS PrivateLinkuma tecnologia que permite acessar o Lake Formation de forma privada APIs sem um gateway de internet, dispositivo NAT, conexão VPN ou conexão. Direct Connect As instâncias em sua VPC não precisam de endereços IP públicos para se comunicar com o Lake Formation. APIs O tráfego entre sua VPC e Lake Formation não sai da rede Amazon.

Cada endpoint de interface é representado por uma ou mais Interfaces de Rede Elástica nas sub-redes.

Para obter mais informações, consulte Endpoints da VPC da interface (AWS PrivateLink) no Manual do Usuário do Amazon VPC.

Considerações sobre endpoints da VPC do Lake Formation

Antes de configurar um endpoint da VPC de interface para o Lake Formation, revise Propriedades e limitações do endpoint de interface no Guia do usuário da Amazon VPC.

O Lake Formation oferece suporte para fazer chamadas para todas as ações de API de sua VPC. Você pode usar o Lake Formation com VPC endpoints em tudo Regiões da AWS que ofereça suporte aos endpoints Lake Formation e Amazon VPC.

Criar um endpoint da VPC de interface para o Lake Formation

Você pode criar um VPC endpoint para o serviço Lake Formation usando o console Amazon VPC ou o (). AWS Command Line Interface AWS CLI Para obter mais informações, consulte Criar um endpoint de interface no Guia do usuário da Amazon VPC.

Crie um endpoint da VPC para o Lake Formation usando o seguinte nome de serviço:

- com.amazonaws. *region*. formação de lago

Se você habilitar o DNS privado para o endpoint, poderá fazer solicitações de API para Lake Formation usando seu nome DNS padrão para a região, por exemplo, `lakeformation.us-east-1.amazonaws.com`.

Para mais informações, consulte Acessar um serviço por um endpoint de interface no Guia do usuário da Amazon VPC.

Criar uma política de endpoint da VPC para o Lake Formation

O Lake Formation oferece suporte a políticas de endpoint da VPC. Uma política de endpoint é uma política baseada em recursos que você anexa a um endpoint VPC para controlar quais AWS entidades principais podem usar o endpoint para acessar um serviço. AWS

Você pode anexar uma política de endpoint ao endpoint da VPC que controla o acesso ao Lake Formation. Essa política especifica as seguintes informações:

- A entidade principal que pode realizar ações.
- As ações que podem ser realizadas.
- Os recursos aos quais as ações podem ser aplicadas.

Para mais informações, consulte Controlar o acesso a serviços com VPC endpoints no Guia do usuário da Amazon VPC.

Exemplo: política de endpoint da VPC para ações do Lake Formation

O exemplo a seguir da política de endpoints da VPC para o Lake Formation permite o fornecimento de credenciais usando as permissões do Lake Formation. Você pode usar essa política para executar consultas usando as permissões do Lake Formation de um cluster do Amazon Redshift ou de Amazon EMR um cluster localizado em uma sub-rede privada.

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

> **Note**
>
> Se você não anexar uma política ao criar um endpoint, uma política padrão que permite acesso total ao serviço será anexada.

Para obter mais informações, consulte estes tópicos na documentação da Amazon VPC:

- O que é Amazon VPC?
- Criar um endpoint de interface
- Usar políticas de endpoint da VPC

AWS Lake Formation tutoriais

Os tutoriais a seguir estão organizados em três faixas e fornecem step-by-step instruções sobre como criar um data lake, ingerir dados, compartilhar e proteger data lakes usando: AWS Lake Formation

1. Criar um data lake e ingerir dados: aprenda a criar um data lake e use esquemas para mover, armazenar, catalogar, limpar e organizar seus dados. Você também aprenderá a configurar tabelas controladas. Uma tabela controlada é um novo tipo de tabela do Amazon S3 que suporta transações atômicas, consistentes, isoladas e duráveis (ACID).

   Antes de começar, certifique-se de ter concluído as etapas em Introdução ao Lake Formation.

   - Criando um data lake a partir de um AWS CloudTrail origem

     Crie e carregue seu primeiro data lake usando seus próprios CloudTrail registros como fonte de dados.

   - Criação de um data lake a partir de uma fonte JDBC no Lake Formation

     Crie um data lake usando um dos seus armazenamentos de dados acessíveis por JDBC, como um banco de dados relacional, como fonte de dados.

2. Proteger data lakes: aprenda a usar controles de acesso baseados em tags e em nível de linha para proteger e gerenciar com eficácia o acesso aos seus data lakes.

   - Como configurar permissões para formatos de armazenamento de tabelas abertas no Lake Formation

     Este tutorial demonstra como configurar permissões para formatos de tabela transacional de código aberto (tabelas do Apache Iceberg, Apache Hudi e Linux Foundation Delta Lake) no Lake Formation.

   - Como gerenciar um data lake usando o controle de acesso baseado em tags do Lake Formation

     Aprenda a gerenciar o acesso aos dados em um data lake usando o controle de acesso baseado em tags no Lake Formation.

   - Como proteger os data lakes com controle de acesso em nível de linha

     Aprenda a configurar permissões em nível de linha que permitem restringir o acesso a linhas específicas com base nas políticas de conformidade e governança de dados no Lake Formation.

3. Compartilhar dados: aprenda a compartilhar seus dados com segurança Contas da AWS usando o controle de acesso baseado em tags (TBAC) e gerencie permissões granulares em conjuntos de dados compartilhados entre eles. Contas da AWS
   - Compartilhamento de um Data Lake usando controle de acesso baseado em tags do Lake Formation e recursos nomeados

     Neste tutorial, você aprenderá como compartilhar seus dados com segurança Contas da AWS usando o Lake Formation.
   - Como compartilhar um data lake usando o controle de acesso refinado do Lake Formation

     Neste tutorial, você aprende como compartilhar conjuntos de dados de forma rápida e fácil usando o Lake Formation ao gerenciar vários Contas da AWS com AWS Organizations.

Tópicos

- Criando um data lake a partir de um AWS CloudTrail origem
- Criação de um data lake a partir de uma fonte JDBC no Lake Formation
- Como configurar permissões para formatos de armazenamento de tabelas abertas no Lake Formation
- Como gerenciar um data lake usando o controle de acesso baseado em tags do Lake Formation
- Como proteger os data lakes com controle de acesso em nível de linha
- Compartilhamento de um Data Lake usando controle de acesso baseado em tags do Lake Formation e recursos nomeados
- Como compartilhar um data lake usando o controle de acesso refinado do Lake Formation

Criando um data lake a partir de um AWS CloudTrail origem

Este tutorial orienta você pelas ações a serem tomadas no console do Lake Formation para criar e carregar seu primeiro data lake a partir de uma AWS CloudTrail fonte.

High-level etapas para criar um data lake

1. Registre um caminho do Amazon Simple Storage Service (Amazon S3) como um data lake.
2. Conceda permissões do Lake Formation para gravar no catálogo de dados e nos locais do Amazon S3 no data lake.
3. Crie um banco de dados para organizar as tabelas de metadados no catálogo de dados.

4. Use um esquema para criar um fluxo de trabalho. Execute o fluxo de trabalho para ingerir dados de uma fonte de dados.
5. Configure suas permissões do Lake Formation para permitir que outras pessoas gerenciem dados no catálogo de dados e no Data Lake.
6. Configure o Amazon Athena para consultar os dados que você importou no seu data lake do Amazon S3.
7. Para alguns tipos de armazenamento de dados, configure o Amazon Redshift Spectrum para consultar os dados que você importou para o seu data lake do Amazon S3.

Tópicos

- Público-alvo
- Pré-requisitos
- Etapa 1: Criar um usuário de analista de dados
- Etapa 2: adicionar permissões para leitura AWS CloudTrail registros para a função de fluxo de trabalho
- Etapa 3: Criar um bucket do Amazon S3 para o data lake
- Etapa 4: Registrar um caminho do Amazon S3
- Etapa 5: Conceder permissões de local de dados
- Etapa 6: Criar um banco de dados no catálogo de dados
- Etapa 7: Conceder permissões de dados
- Etapa 8: Usar um esquema para criar um fluxo de trabalho
- Etapa 9: Executar o fluxo de trabalho
- Etapa 10: Conceder SELECT nas tabelas
- Etapa 11: consulte o data lake usando Amazon Athena

Público-alvo

A tabela a seguir lista as perfis usadas neste tutorial para criar um data lake.

Público-alvo

Perfil	Description
Administrador do IAM	Tem a política AWS gerenciada:`AdministratorAccess` . Criar perfis do IAM e buckets do Amazon S3.
Administrador do data lake	Usuário que pode acessar o catálogo de dados, criar bancos de dados e conceder permissões do Lake Formation a outros usuários. Tem menos permissões do IAM do que o administrador do IAM, mas o suficiente para administrar o data lake.
Analista de dados	Usuário que pode executar consultas no data lake. Tem permissões suficientes apenas para executar consultas.
Função de fluxo de trabalho	Perfil com as políticas de IAM necessárias para executar um fluxo de trabalho. Para obter mais informações, consulte (Opcional) Criar um perfil do IAM para fluxos de trabalho.

Pré-requisitos

Antes de começar

- Verifique se você concluiu as tarefas no Configurar AWS Lake Formation.
- Conheça a localização dos seus CloudTrail registros.
- O Athena exige que a pessoa do analista de dados crie um bucket do Amazon S3 para armazenar os resultados da consulta antes de usar o Athena.

Presume-se familiaridade com AWS Identity and Access Management (IAM). Para obter informações sobre o IAM, consulte o Guia do usuário do IAM.

Etapa 1: Criar um usuário de analista de dados

Esse usuário tem as permissões de privilégio mínimo para consultar o data lake.

1. Abra o console do IAM em https://console.aws.amazon.com/iam. Faça login como usuário com a política `AdministratorAccess` AWS gerenciada.
2. Crie um usuário chamado `datalake_user` com as seguintes configurações:
   - Habilite Console de gerenciamento da AWS o acesso.
   - Defina uma senha e não solicite redefinição de senha.
   - Anexe a política `AmazonAthenaFullAccess` AWS gerenciada.
   - Anexe a seguinte política em linha. Atribua o nome `DatalakeUserBasic` à política.

```
{
    "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": "*"
        }
    ]
}
```

Etapa 2: adicionar permissões para leitura AWS CloudTrail registros para a função de fluxo de trabalho

1. Anexe a política em linha a seguir ao perfil `LakeFormationWorkflowRole`. A política concede permissão para ler seus AWS CloudTrail registros. Atribua o nome `DatalakeGetCloudTrail` à política.

   Para criar a função `LakeFormationWorkflowRole`, consulte (Opcional) Criar um perfil do IAM para fluxos de trabalho.

   > **Important**
   >
   > *<your-s3-cloudtrail-bucket>*Substitua pela localização dos seus CloudTrail dados no Amazon S3.

   JSON

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

2. Verifique se há três políticas vinculadas ao perfil.

Etapa 3: Criar um bucket do Amazon S3 para o data lake

Crie o bucket do Amazon S3 que será o local raiz do seu data lake.

1. Abra o console do Amazon S3 em https://console.aws.amazon.com/s3/e faça login como usuário com acesso administrativo.

2. Escolha Create bucket e acesse o assistente para criar um bucket chamado`<yourName>-datalake-cloudtrail`, onde `<yourName>` está seu primeiro nome e sobrenome. Por exemplo: `jdoe-datalake-cloudtrail`.

   Para obter instruções detalhadas sobre como criar um bucket do Amazon S3, consulte Como criar um bucket.

Etapa 4: Registrar um caminho do Amazon S3

Registre um caminho do Amazon S3 como o local raiz do seu data lake.

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/. Faça login como administrador de data lake.
2. No painel de navegação, em Registrar e ingerir, escolha Locais do data lake.
3. Selecione Registrar local e, em seguida, Navegar.
4. Selecione o bucket do `<yourName>-datalake-cloudtrail` que você criou anteriormente, aceite o perfil padrão do IAM `AWSServiceRoleForLakeFormationDataAccess` e selecione Registrar local.

   Para obter mais informações sobre o registro de locais, consulte Adicionar uma localização do Amazon S3 ao seu data lake.

Etapa 5: Conceder permissões de local de dados

As entidades principais devem ter permissões de local de dados em um local de data lake para criar tabelas ou bancos de dados do catálogo de dados que apontem para esse local. Você deve conceder permissões de local de dados ao perfil do IAM para fluxos de trabalho para que o fluxo de trabalho possa gravar no destino da ingestão de dados.

1. No painel de navegação, em Permissões, selecione Locais de dados.
2. Selecione Conceder e, na caixa de diálogo Conceder permissões, faça estas seleções:

   a. Em Usuários e perfis do IAM, escolha `LakeFormationWorkflowRole`.

   b. Para locais de armazenamento, escolha seu bucket `<yourName>-datalake-cloudtrail`.

3. Selecione Conceder.

Para obter mais informações sobre permissões de local de dados, consulte Underlying data access control.

Etapa 6: Criar um banco de dados no catálogo de dados

As tabelas de metadados no catálogo de dados do Lake Formation são armazenadas em um banco de dados.

1. No painel de navegação, em catálogo de dados, escolha Bancos de dados.
2. selecione Criar banco de dados e, em Informações do banco de dados, digite o nome `lakeformation_cloudtrail`.
3. Deixe os outros campos em branco e escolha Criar banco de dados.

Etapa 7: Conceder permissões de dados

Você deve conceder permissões para criar tabelas de metadados no catálogo de dados. Como o fluxo de trabalho será executado com o perfil de `LakeFormationWorkflowRole`, você deve conceder essas permissões ao perfil.

1. No console do Lake Formation, no painel de navegação, em catálogo de dados, selecione Bancos de dados.
2. Selecione o banco de dados `lakeformation_cloudtrail`, e na lista suspensa Ações, selecione Conceder sob o título Permissões.
3. Na caixa de diálogo Conceder permissões de dados, faça estas seleções:
    a. Em Entidades principais, em Usuário e perfis do IAM, escolha `LakeFormationWorkflowRole`.
    b. Em Recursos do catálogo LF-Tags ou, escolha Recursos do Catálogo de Dados Nomeados.
    c. Para Bancos de dados, você deve ver que o banco de dados `lakeformation_cloudtrail` já foi adicionado.
    d. Em Permissões do banco de dados, selecione Criar tabela, Alterar e Eliminar, e desmarque Super se estiver selecionado.
4. Selecione Conceder.

Para obter mais informações sobre como conceder permissões ao Lake Formation, consulte Gerenciando permissões do Lake Formation.

Etapa 8: Usar um esquema para criar um fluxo de trabalho

Para ler os CloudTrail registros, entender sua estrutura e criar as tabelas apropriadas no Catálogo de Dados, precisamos configurar um fluxo de trabalho que consiste em AWS Glue rastreadores, tarefas, acionadores e fluxos de trabalho. Os esquemas do Lake Formation simplificam esse processo.

O fluxo de trabalho gera trabalhos, crawlers e gatilhos que descobrem e ingerem dados em seu data lake. Você cria um fluxo de trabalho com base em um dos esquemas predefinidos do Lake Formation.

1. No console do Lake Formation, no painel de navegação, selecione Esquemas em Ingestão e, depois, selecione Usar esquema.
2. Na página Usar um blueprint, em Tipo de blueprint, escolha. AWS CloudTrail
3. Em Importar fonte, escolha uma CloudTrail fonte e uma data de início.
4. Em Destino de importação, especifique estes parâmetros:

Bancos de dados de destino	`lakeformation_cloudtrail`
Local de armazenamento de destino	`s3://<yourName> -datalake-cloudtrail`
Formato de dados	Parquet

5. Para frequência de importação, selecione Executar sob demanda.
6. Em Opções de importação, especifique estes parâmetros:

Nome do fluxo de trabalho	`lakeformationcloudtrailtest`
Perfil do IAM	`LakeFormationWorkflowRole`
Prefixo da tabela	`cloudtrailtest` **Note** Deve estar em letras minúsculas.

7. Escolha Criar e aguarde até que o console informe que o fluxo de trabalho foi criado com sucesso.

> **Tip**
>
> Você recebeu a seguinte mensagem de erro?
> `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...`
> Nesse caso, verifique se você substituiu a política *<account-id>* em linha do usuário administrador do data lake por um número de AWS conta válido.

Etapa 9: Executar o fluxo de trabalho

Como você especificou que o fluxo de trabalho é executado sob demanda, você deve iniciá-lo manualmente.

- Na página Esquemas, selecione o fluxo de trabalho `lakeformationcloudtrailtest` e, no menu Ações, selecione Iniciar.

  À medida que o fluxo de trabalho é executado, você pode ver seu progresso na coluna de Status da última execução. Escolha o botão de atualização ocasionalmente.

  O status vai de EM EXECUÇÃO, para Descoberta, para Importação, e CONCLUÍDO.

  Quando o fluxo de trabalho for concluído:

  - O catálogo de dados terá novas tabelas de metadados.
  - Seus CloudTrail registros serão ingeridos no data lake.

  Se o fluxo de trabalho falhar, faça o seguinte:

  a. Selecione o fluxo de trabalho e, no menu Ações, selecione Exibir gráfico.

     O fluxo de trabalho é aberto no console do AWS Glue.

  b. Certifique-se de que o fluxo de trabalho esteja selecionado e acesse a guia Histórico.

  c. Em Histórico, selecione a execução mais recente e selecione Exibir informações da execução.

d. Selecione um trabalho ou crawler com falha no gráfico dinâmico (runtime) e revise a mensagem de erro. Os nós com falha são vermelhos ou amarelos.

Etapa 10: Conceder SELECT nas tabelas

Você deve conceder a permissão `SELECT` nas novas tabelas do catálogo de dados para que o analista de dados possa consultar os dados para os quais as tabelas apontam.

> **Note**
>
> Um fluxo de trabalho concede automaticamente a permissão `SELECT` nas tabelas que ele cria ao usuário que o executou. Como o administrador do data lake executou esse fluxo de trabalho, você deve conceder `SELECT` ao analista de dados.

1. No console do Lake Formation, no painel de navegação, em catálogo de dados, selecione Bancos de dados.
2. Selecione o banco de dados `lakeformation_cloudtrail`, e na lista suspensa Ações, selecione Conceder sob o título Permissões.
3. Na caixa de diálogo Conceder permissões de dados, faça estas seleções:
   a. Em Entidades principais, em Usuário e perfis do IAM, escolha `datalake_user`.
   b. Em Recursos do catálogo LF-Tags ou, escolha Recursos do catálogo de dados nomeados.
   c. Para Bancos de dados, o banco de dados `lakeformation_cloudtrail` já deve estar selecionado.
   d. Para Tabelas, selecione `cloudtrailtest-cloudtrail`.
   e. Em Permissões de tabela e coluna, clique em Selecionar.
4. Selecione Conceder.

A próxima etapa é executada como analista de dados.

Etapa 11: consulte o data lake usando Amazon Athena

Use o Amazon Athena console para consultar os CloudTrail dados em seu data lake.

1. Abra o console do Athena em https://console.aws.amazon.com/athena/e faça login como analista de dados, usuário. `datalake_user`
2. Se necessário, escolha Começar para continuar com o editor de consultas do Athena.
3. Em Data source (Fonte de dados), selecione AwsDataCatalog.
4. Para o Banco de dados, selecione `lakeformation_cloudtrail`.

   A lista de Tabelas é preenchida.
5. No menu flutuante (3 pontos horizontais) ao lado da tabela `cloudtrailtest-cloudtrail`, selecione Exibir tabela e selecione Executar.

   A consulta é executada e exibe 10 linhas de dados.

   Se você nunca usou o Athena antes, primeiro deve configurar um local do Amazon S3 no console do Athena para armazenar os resultados da consulta. O `datalake_user` deve ter as permissões exigidas para acessar o bucket do Amazon S3 escolhido.

> **Note**
>
> Agora que você concluiu o tutorial, conceda permissões de dados e permissões de local de dados às entidades principais da sua organização.

Criação de um data lake a partir de uma fonte JDBC no Lake Formation

Este tutorial orienta você pelas etapas a serem seguidas no AWS Lake Formation console para criar e carregar seu primeiro data lake a partir de uma fonte JDBC usando o Lake Formation.

Tópicos

- Público-alvo
- Pré-requisitos do tutorial JDBC
- Etapa 1: Criar um usuário de analista de dados
- Etapa 2: criar uma conexão no AWS Glue
- Etapa 3: Criar um bucket do Amazon S3 para o data lake
- Etapa 4: Registrar um caminho do Amazon S3

- Etapa 5: Conceder permissões de local de dados
- Etapa 6: Criar um banco de dados no catálogo de dados
- Etapa 7: Conceder permissões de dados
- Etapa 8: Usar um esquema para criar um fluxo de trabalho
- Etapa 9: Executar o fluxo de trabalho
- Etapa 10: Conceder SELECT nas tabelas
- Etapa 11: consulte o data lake usando Amazon Athena
- Etapa 12: Consulte os dados no data lake usando o Amazon Redshift Spectrum
- Etapa 13: Conceder ou revogar permissões do Lake Formation usando o Amazon Redshift Spectrum

Público-alvo

A tabela a seguir lista as funções que são usadas neste tutorial do JDBC AWS Lake Formation.

Função	Description
Administrador do IAM	Um usuário que pode criar AWS Identity and Access Management (IAM) usuários e funções e buckets do Amazon Simple Storage Service (Amazon S3). Tem a política `AdministratorAccess` AWS gerenciada.
Administrador do data lake	Um usuário que pode acessar o catálogo de dados, criar bancos de dados e conceder permissões do Lake Formation a outros usuários. Possui menos permissões do IAM do que o administrador do IAM, mas o suficiente para administrar o data lake.
Analista de dados	Um usuário que pode executar consultas no data lake. Tem permissões suficientes apenas para executar consultas.

Função	Description
Função do fluxo de trabalho	Um perfil com as políticas de IAM necessárias para executar um fluxo de trabalho.

Para obter informações sobre os pré-requisitos para concluir o tutorial, consulte Pré-requisitos do tutorial JDBC.

Pré-requisitos do tutorial JDBC

Antes de começar o tutorial do AWS Lake Formation JDBC, certifique-se de ter feito o seguinte:

- Conclua as tarefas em Introdução ao Lake Formation.
- Escolha um armazenamento JDBC-accessible de dados que você deseja usar para o tutorial.
- Reúna as informações necessárias para criar uma conexão AWS Glue do tipo JDBC. Esse objeto do Catálogo de Dados inclui a URL do armazenamento de dados, as credenciais de login e, se o armazenamento de dados foi criado em uma Amazon Virtual Private Cloud (Amazon VPC), VPC-specific informações adicionais de configuração. Para obter mais informações, consulte Definir conexões no catálogo de dados AWS Glue no Guia do desenvolvedor do AWS Glue .

O tutorial pressupõe que você esteja familiarizado com AWS Identity and Access Management (IAM). Para obter informações sobre o IAM, consulte o Guia do usuário do IAM.

Para começar, vá para the section called “Etapa 1: Criar um usuário de analista de dados”.

Etapa 1: Criar um usuário de analista de dados

Nesta etapa, você cria um usuário AWS Identity and Access Management (IAM) para ser o analista de dados do seu data lake em AWS Lake Formation.

Esse usuário tem o conjunto mínimo de permissões para consultar o data lake.

1. Abra o console do IAM em https://console.aws.amazon.com/iam. Faça login como usuário com a política `AdministratorAccess` AWS gerenciada.
2. Crie um usuário chamado `datalake_user` com as seguintes configurações:
   - Habilite Console de gerenciamento da AWS o acesso.

- Defina uma senha e não solicite redefinição de senha.
- Anexe a política `AmazonAthenaFullAccess` AWS gerenciada.
- Anexe a seguinte política em linha. Atribua o nome `DatalakeUserBasic` à política.

```
{
    "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": "*"
        }
    ]
}
```

Etapa 2: criar uma conexão no AWS Glue

> **Note**
>
> Ignore esta etapa se você já tiver uma conexão AWS Glue com sua fonte de dados JDBC.

AWS Lake Formation acessa fontes de dados JDBC por meio de uma conexão. AWS Glue Uma conexão é um objeto do catálogo de dados que contém todas as informações necessárias para se conectar à fonte de dados. Você pode criar uma conexão usando o console do AWS Glue.

Para criar uma conexão

1. Abra o AWS Glue console em https://console.aws.amazon.com/glue/ e faça login como usuário com acesso administrativo.
2. No painel de navegação, em Data catalog (catálogo de dados), selecione Connections (Conexões).
3. Na página Conectores, escolha Criar conexão.
4. Na página Escolher fonte de dados, escolha JDBC como o tipo de conexão. Escolha Próximo.
5. Continue através do assistente de conexão e salve a conexão.

   Para obter informações sobre como criar uma conexão, consulte Propriedades da conexão JDBC do AWS Glue no Guia do desenvolvedor do AWS Glue .

Etapa 3: Criar um bucket do Amazon S3 para o data lake

Nesta etapa, você criará o bucket do Amazon Simple Storage Service (Amazon S3) que será a localização raiz do data lake.

1. Abra o console do Amazon S3 em https://console.aws.amazon.com/s3/e faça login como usuário com acesso administrativo.
2. Escolha Create bucket e acesse o assistente para criar um bucket chamado`<yourName>-datalake-tutorial`, onde `<yourName>` está seu primeiro nome e sobrenome. Por exemplo: `jdoe-datalake-tutorial`.

   Para obter instruções detalhadas sobre como criar um bucket do Amazon S3, consulte Como criar um bucket do S3? no Guia do usuário do Amazon Simple Storage Service.

Etapa 4: Registrar um caminho do Amazon S3

Nesta etapa, você registra um caminho do Amazon Simple Storage Service (Amazon S3) como a localização raiz do data lake.

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/. Faça login como administrador de data lake.
2. No painel de navegação, em Administração em Locais de data lake.
3. Escolha Registrar localização e, em seguida, escolha Procurar.

4. Selecione o bucket *<yourName>*`-datalake-tutorial` que você criou anteriormente, aceite o perfil padrão do IAM `AWSServiceRoleForLakeFormationDataAccess` e escolha Registrar localização.

   Para obter mais informações sobre o registro de locais, consulte Adicionar uma localização do Amazon S3 ao seu data lake.

Etapa 5: Conceder permissões de local de dados

As entidades principais devem ter permissões de local de dados em um local de data lake para criar tabelas ou bancos de dados do catálogo de dados que apontem para esse local. Você deve conceder permissões de local de dados ao perfil do IAM para fluxos de trabalho para que o fluxo de trabalho possa gravar no destino da ingestão de dados.

1. No console do Lake Formation, no painel de navegação, em Permissões, escolha Localizações de dados.
2. Escolha Conceder e, na caixa de diálogo Conceder permissões, faça o seguinte:
   a. Em Usuários e perfis do IAM, escolha `LakeFormationWorkflowRole`.
   b. Para locais de armazenamento, escolha seu bucket *<yourName>*`-datalake-tutorial`.
3. Selecione Conceder.

Para obter mais informações sobre permissões de local de dados, consulte Underlying data access control.

Etapa 6: Criar um banco de dados no catálogo de dados

As tabelas de metadados no catálogo de dados do Lake Formation são armazenadas em um banco de dados.

1. No console do Lake Formation, no painel de navegação, em catálogo de dados, escolha Bancos de dados.
2. Escolha Criar banco de dados e, em Detalhes do banco de dados, insira o nome `lakeformation_tutorial`.
3. Deixe os outros campos em branco e escolha Criar banco de dados.

Etapa 7: Conceder permissões de dados

Você deve conceder permissões para criar tabelas de metadados no catálogo de dados. Como o fluxo de trabalho é executado com a função `LakeFormationWorkflowRole`, você deve conceder essas permissões à função.

1. No console do Lake Formation, no painel de navegação, em Permissões, escolha Permissões do data lake.
2. Escolha Conceder e, na caixa de diálogo Conceder permissões de dados, faça o seguinte:
   a. Em Entidades principais, em Usuário e perfis do IAM, escolha `LakeFormationWorkflowRole`.
   b. Em Recursos do catálogo LF-Tags ou, escolha Recursos do catálogo de dados nomeados.
   c. Em Bancos de dados, escolha o banco de dados que você criou anteriormente, `lakeformation_tutorial`.
   d. Em Permissões do banco de dados, selecione Criar tabela, Alterar e Eliminar, e desmarque Super se estiver selecionado.
3. Selecione Conceder.

Para obter mais informações sobre como conceder permissões ao Lake Formation, consulte Visão geral das permissões do Lake Formation.

Etapa 8: Usar um esquema para criar um fluxo de trabalho

O AWS Lake Formation fluxo de trabalho gera AWS Glue trabalhos, rastreadores e gatilhos que descobrem e ingerem dados em seu data lake. Você cria um fluxo de trabalho com base em um dos esquemas predefinidos do Lake Formation.

1. No console do Lake Formation, no painel de navegação, escolha Esquemas e, em seguida, escolha Usar esquema.
2. Na página Usar um esquema, em Tipo de esquema, escolha Snapshot de banco de dados.
3. Em Fonte de importação, em Conexão de banco de dados, escolha a conexão que você acabou de criar, `datalake-tutorial` ou escolha uma conexão existente para sua fonte de dados.
4. Em Caminho de dados de origem, insira o caminho do qual ingerir dados, no formulário *<database>*/*<schema>*/*<table>*.

Você pode substituir o curinga percentual (%) por esquema ou tabela. Para bancos de dados que oferecem suporte a esquemas, insira*<database>*/*<schema>*/% para corresponder a todas as tabelas contidas*<schema>*. *<database>* O Oracle Database e o MySQL não oferecem suporte ao esquema no caminho; em vez disso, insira /%. *<database>* Para o Oracle Database, *<database>* é o identificador do sistema (SID).

Por exemplo, se um banco de dados Oracle tiver `orcl` como SID, insira `orcl/%` para corresponder a todas as tabelas às quais o usuário especificado na conexão JDCB tem acesso.

> **Important**
>
> Este campo diferencia letras maiúsculas de minúsculas.

5. Em Destino de importação, especifique estes parâmetros:

Bancos de dados de destino	`lakeformation_tutorial`
Local de armazenamento de destino	`s3://`*<yourName>* `-datalake-tutorial`
Formato de dados	(Escolha Parquet ou CSV)

6. Para frequência de importação, escolha Executar sob demanda.
7. Em Opções de importação, especifique estes parâmetros:

Nome do fluxo de trabalho	`lakeformationjdbctest`
Perfil do IAM	`LakeFormationWorkflowRole`
Prefixo da tabela	`jdbctest` **Note** Deve estar em letras minúsculas.

8. Escolha Criar e aguarde até que o console informe que o fluxo de trabalho foi criado com sucesso.

> **Tip**
>
> Você recebeu a seguinte mensagem de erro?
> `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...`
> Nesse caso, verifique se você substituiu a política *<account-id>* em linha do usuário administrador do data lake por um número de AWS conta válido.

Etapa 9: Executar o fluxo de trabalho

Como você especificou que o fluxo de trabalho é executado sob demanda, você deve iniciá-lo manualmente em. AWS Lake Formation

1. No console do Lake Formation, na página Esquemas, selecione o fluxo de trabalho `lakeformationjdbctest`.
2. Escolha Ações e, em seguida, escolha Iniciar.
3. À medida que o fluxo de trabalho é executado, visualize seu progresso na coluna Status da última execução. Escolha o botão de atualização ocasionalmente.

   O status vai de EM EXECUÇÃO para Descoberta, para Importação e para CONCLUÍDO.

   Quando o fluxo de trabalho for concluído:

   - O catálogo de dados tem novas tabelas de metadados.
   - Seus dados são ingeridos no data lake.

   Se o fluxo de trabalho falhar, faça o seguinte:

   a. Selecione o fluxo de trabalho. Escolha Ações e Exibir gráfico.

      O fluxo de trabalho é aberto no console do AWS Glue.

   b. Selecione o fluxo de trabalho e escolha a guia Histórico.

   c. Selecione a execução mais recente e escolha Exibir detalhes da execução.

d. Selecione um trabalho ou crawler com falha no gráfico dinâmico (runtime) e revise a mensagem de erro. Os nós com falha são vermelhos ou amarelos.

Etapa 10: Conceder SELECT nas tabelas

Você deve conceder a `SELECT` permissão nas novas tabelas do Catálogo de Dados AWS Lake Formation para que o analista de dados possa consultar os dados para os quais as tabelas apontam.

> **Note**
>
> Um fluxo de trabalho concede automaticamente a permissão `SELECT` nas tabelas que ele cria ao usuário que o executou. Como o administrador do data lake executou esse fluxo de trabalho, você deve conceder `SELECT` ao analista de dados.

1. No console do Lake Formation, no painel de navegação, em Permissões, escolha Permissões do data lake.
2. Escolha Conceder e, na caixa de diálogo Conceder permissões de dados, faça o seguinte:
   a. Em Entidades principais, em Usuário e perfis do IAM, escolha `datalake_user`.
   b. Em Recursos do catálogo LF-Tags ou, escolha Recursos do catálogo de dados nomeados.
   c. Para Bancos de dados, escolha `lakeformation_tutorial`.

      A lista de Tabelas é preenchida.
   d. Para Tabelas, selecione uma ou mais tabelas da fonte de dados.
   e. Em Permissões de tabela e coluna, escolha Selecionar.
3. Selecione Conceder.

A próxima etapa é executada como analista de dados.

Etapa 11: consulte o data lake usando Amazon Athena

Use o Amazon Athena console para consultar os dados em seu data lake.

1. Abra o console do Athena em https://console.aws.amazon.com/athena/e faça login como analista de dados, usuário. `datalake_user`

2. Se necessário, escolha Começar para continuar com o editor de consultas do Athena.
3. Em Data source (Fonte de dados), selecione AwsDataCatalog.
4. Para o Banco de dados, selecione `lakeformation_tutorial`.

   A lista Tabelas é preenchida.
5. No menu pop-up ao lado de uma das tabelas, escolha Visualizar tabela.

   A consulta é executada e exibe 10 linhas de dados.

Etapa 12: Consulte os dados no data lake usando o Amazon Redshift Spectrum

Você pode configurar o Amazon Redshift Spectrum para consultar os dados importados para o data lake do Amazon Simple Storage Service (Amazon S3). Primeiro, crie uma função AWS Identity and Access Management (IAM) que seja usada para iniciar o cluster Amazon Redshift e consultar os dados do Amazon S3. Em seguida, conceda a essa função as permissões `Select` nas tabelas que você deseja consultar. Em seguida, conceder permissões do usuário para usar o Editor de consultas do Amazon Redshift. Por fim, crie um cluster do Amazon Redshift e execute consultas.

Você cria o cluster como administrador e consulta o cluster como analista de dados.

Para obter mais informações sobre o Amazon Redshift Spectrum, consulte Usar o Amazon Redshift Spectrum para consultar dados externos no Guia do desenvolvedor de banco de dados do Amazon Redshift.

Para configurar permissões para executar consultas do Amazon Redshift

1. Abra o console do IAM em https://console.aws.amazon.com/iam/. Faça login como usuário com a política `AdministratorAccess` AWS gerenciada.
2. No painel de navegação, selecione Políticas.

   Se essa for a primeira vez que você escolhe Políticas, a página Bem-vindo às políticas gerenciadas será exibida. Escolha Começar.
3. Escolha Criar política.
4. Escolha a guia JSON.
5. Cole no seguinte documento de política JSON.

```
{
    "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. Ao terminar, escolha Review (Revisar) para revisar a política. O validador de política indica se há qualquer erro de sintaxe.
7. Na página Revisar política, insira o Nome como **`RedshiftLakeFormationPolicy`** na política que você está criando. Insira uma Description (Descrição) (opcional). Revise o Resumo da política para ver as permissões que são concedidas pela política. Em seguida, escolha Criar política para salvar seu trabalho.
8. No painel de navegação do console do IAM, escolha Roles (Perfis) e, em seguida, Create role (Criar perfil).
9. Em Select trusted entity (Selecionar entidade confiável), escolha AWS Service (Serviço).
10. Escolha o serviço do Amazon Redshift para assumir essa função.
11. Escolha o caso de uso Redshift Customizable (Personalizável pelo Redshift) para o serviço. Então, escolha Próximo: permissões.
12. Procure a política de permissões que você criou, `RedshiftLakeFormationPolicy` e marque a caixa de seleção ao lado do nome da política na lista.

13. Escolha Próximo: tags.
14. Selecione Próximo: revisar.
15. Em Role name (Nome da função), insira o nome **`RedshiftLakeFormationRole`**.
16. (Opcional) Em Descrição do perfil, insira uma descrição para o novo perfil.
17. Reveja a função e escolha Criar função.

Para conceder permissões **`Select`** na tabela a ser consultada no banco de dados do Lake Formation

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/. Faça login como administrador de data lake.
2. No painel de navegação, em Permissões, escolha Permissões do Data Lake e, em seguida, escolha Conceder.
3. Forneça as informações a seguir:
   - Para usuários e funções do IAM, selecione o perfil do IAM criado, `RedshiftLakeFormationRole`. Ao executar o editor de consulta do Amazon Redshift, ele usa esse perfil do IAM para conceder permissão aos dados.
   - Para o Banco de dados, selecione `lakeformation_tutorial`.

     A lista de tabelas é preenchida.
   - Em Tabela, escolha uma tabela na fonte de dados a ser consultada.
   - Escolha a permissão Selecionar tabela.
4. Selecione Conceder.

Para configurar o Amazon Redshift Spectrum e executar consultas

1. Abra o console do Amazon Redshift em https://console.aws.amazon.com/redshift. Faça login como usuário `Administrator`.
2. Selecione Criar cluster.
3. Na página Criar cluster, insira `redshift-lakeformation-demo` o Identificador do cluster.
4. Para o Tipo de nó, selecione dc2.large.
5. Role para baixo e, em Configurações do banco de dados, insira ou aceite estes parâmetros:
   - Nome de usuário administrador: `awsuser`

- Senha do usuário administrador: (*Choose a password*)

6. Expanda as permissões do cluster e, em Funções do IAM disponíveis, escolha RedshiftLakeFormationRole. Depois, escolha Add IAM role (Adicionar perfil do IAM).
7. Se você precisar usar uma porta diferente do valor padrão de 5439, ao lado de Configurações adicionais, desative a opção Usar padrões. Expanda a seção Configurações do banco de dados e, em seguida, insira um novo número de Porta do banco de dados.
8. Selecione Criar cluster.

   A página Clusters é carregada.
9. Espere até que o status do cluster se torne Disponível. Escolha o ícone de atualização periodicamente.
10. Conceda permissão ao analista de dados para executar consultas no cluster. Para fazer isso, execute as etapas a seguir:

    a. Abra o console do IAM em https://console.aws.amazon.com/iam/e faça login como `Administrator` usuário.

    b. No painel de navegação, escolha Usuários e anexe as seguintes políticas gerenciadas ao usuário `datalake_user`.

    - `AmazonRedshiftQueryEditor`
    - `AmazonRedshiftReadOnlyAccess`
11. Saia do console do Amazon Redshift e entre novamente como usuário `datalake_user`.
12. Na barra de ferramentas vertical esquerda, escolha o ícone EDITOR para abrir o editor de consultas e conectar-se ao cluster. Se a caixa de diálogo Conectar ao banco de dados for exibida, escolha o nome do cluster `redshift-lakeformation-demo` e insira o nome do banco de dados **`dev`**, o nome do usuário **`awsuser`** e a senha que você criou. Então escolha Connect to database (Conectar-se ao banco de dados).

> **Note**
>
> Se os parâmetros de conexão não forem solicitados e outro cluster já estiver selecionado no editor de consultas, escolha Alterar conexão para abrir a caixa de diálogo Conectar ao banco de dados.

13. Na caixa de texto Nova consulta 1, insira e execute a seguinte instrução para mapear o banco de dados `lakeformation_tutorial` no Lake Formation para o nome do esquema Amazon Redshift `redshift_jdbc`:

> **Important**
>
> *<account-id>*Substitua por um número de AWS conta válido e *<region>* por um nome de AWS região válido (por exemplo,`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. Na lista de esquemas, em Selecionar esquema, escolha redshift_jdbc.

    A lista de tabelas é preenchida. O editor de consultas mostra somente as tabelas nas quais você recebeu permissões de data lake do Lake Formation.

15. No menu pop-up ao lado do nome da tabela, escolha Visualizar dados.

    O Amazon Redshift retorna as 10 primeiras linhas.

    Agora você pode executar consultas nas tabelas e colunas para as quais você tem permissões.

Etapa 13: Conceder ou revogar permissões do Lake Formation usando o Amazon Redshift Spectrum

O Amazon Redshift oferece suporte à capacidade de conceder e revogar permissões do Lake Formation em bancos de dados e tabelas usando instruções SQL modificadas. Essas declarações são semelhantes às declarações existentes do Amazon Redshift. Para obter mais informações, consulte CONCEDER e REVOGAR no Guia do desenvolvedor de banco de dados do Amazon Redshift.

Como configurar permissões para formatos de armazenamento de tabelas abertas no Lake Formation

AWS Lake Formation suporta o gerenciamento de permissões de acesso para Open Table Formats (OTFs), como Apache Iceberg, Apache Hudi e Linux Foundation Delta Lake. Neste tutorial, você aprenderá como criar Iceberg, Hudi e Delta Lake com tabelas de manifesto de links simbólicos AWS Glue Data Catalog usando AWS Glue, configurar permissões refinadas usando o Lake Formation e consultar dados usando o Amazon Athena.

> **Note**
>
> AWS os serviços de análise não oferecem suporte a todos os formatos de tabela transacional. Para obter mais informações, consulte Trabalhando com outros AWS serviços. Este tutorial aborda a criação manual de um banco de dados e de uma tabela no Catálogo de Dados usando somente trabalhos do AWS Glue .

Este tutorial inclui um AWS CloudFormation modelo para configuração rápida. É possível revisá-lo e personalizá-lo para atender às suas necessidades.

Tópicos

- Público-alvo
- Pré-requisitos
- Etapa 1: Provisionar os recursos
- Etapa 2: Configurar permissões para uma tabela do Iceberg
- Etapa 3: Configurar permissões para uma tabela do Hudi
- Etapa 4: Configurar permissões para uma tabela do Delta Lake
- Etapa 5: limpar AWS recursos

Público-alvo

Este tutorial é destinado a administradores de IAM, administradores de data lake e analistas de negócios. A tabela a seguir lista os perfis usados neste tutorial para criar uma tabela controlada usando o Lake Formation.

Perfil	Description
Administrador do IAM	Um usuário que pode criar Usuários e perfis do IAM e buckets do Amazon S3. Tem a política `AdministratorAccess` AWS gerenciada.
Administrador do data lake	Um usuário que pode acessar o catálogo de dados, criar bancos de dados e conceder permissões do Lake Formation a outros usuários. Tem menos permissões do IAM do que o administrador do IAM, mas o suficiente para administrar o data lake.
Analista de negócios	Um usuário que pode executar consultas no data lake. Tem permissões para executar consultas.

Pré-requisitos

Antes de começar este tutorial, você deve ter um Conta da AWS que possa fazer login como usuário com as permissões corretas. Para obter mais informações, consulte Inscreva-se em um Conta da AWS.

O tutorial pressupõe que você esteja familiarizado com os perfis e políticas do IAM. Para obter informações sobre o IAM, consulte o Guia do usuário do IAM.

Você precisa configurar os seguintes AWS recursos para concluir este tutorial:

- Usuário administrador do data lake
- Configurações do data lake Formation
- Versão 3 do mecanismo Amazon Athena

Como criar um administrador de data lake

1. Faça login no console do Lake Formation https://console.aws.amazon.com/lakeformation/como usuário administrador. Neste tutorial, você criará recursos na região Leste dos EUA (Norte da Virgínia).

2. No console do Lake Formation, no painel de navegação, em Permissões, selecione Perfis e tarefas administrativas.
3. Selecione Escolher administradores em Administradores de data lake.
4. Na janela Gerenciar administradores do data lake, em Usuários e perfis do IAM, selecione Usuário Administrador do IAM.
5. Escolha Salvar.

Como ativar as configurações do data lake

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/. No painel de navegação, em catálogo de dados, selecione Configurações. Desmarque o seguinte:
   - Use somente o controle de acesso do IAM para novos bancos de dados.
   - Use somente o controle de acesso do IAM para novas tabelas em novos bancos de dados.
2. Em Configurações da versão entre contas, selecione Versão 3 como a versão entre contas.
3. Escolha Salvar.

Como atualizar o mecanismo Amazon Athena para a versão 3

1. Abra o console Athena em. https://console.aws.amazon.com/athena/
2. Selecione o Grupo de trabalho e selecione o grupo de trabalho principal.
3. Certifique-se de que o grupo de trabalho tenha pelo menos a versão 3. Se não estiver, edite o grupo de trabalho, selecione Manual for Atualizar o mecanismo de consulta e selecione a versão 3.
4. Selecione Salvar alterações.

Etapa 1: Provisionar os recursos

Esta seção mostra como configurar os AWS recursos usando um CloudFormation modelo.

Para criar seus recursos usando CloudFormation modelo

1. Faça login no AWS CloudFormation console https://console.aws.amazon.com/cloudformationcomo administrador do IAM na região Leste dos EUA (Norte da Virgínia).
2. Selecione Iniciar Pilha.

3. Na página Criar pilha, selecione Avançar.
4. Digite um Nome de pilha.
5. Selecione Avançar.
6. Na página seguinte, selecione Avançar.
7. Analise os detalhes na página final e selecione Eu reconheço que isso AWS CloudFormation pode criar recursos do IAM.
8. Escolha Criar.

   A criação da pilha pode levar até dois minutos.

Ao iniciar a pilha de Cloud Formation, os seguintes recursos são criados:

- lf-otf-datalake-123456789012 — Bucket do Amazon S3 para armazenar dados

> **Note**
>
> O ID da conta anexado ao nome do bucket do Amazon S3 é substituído pelo ID da sua conta.

- lf-otf-tutorial-123456789012 — Bucket Amazon S3 para armazenar resultados de consultas e scripts de trabalho AWS Glue
- lficebergdb — Banco de dados Iceberg AWS Glue
- lfhudidb — Banco de dados Hudi AWS Glue
- lfdeltadb — Banco de dados Delta AWS Glue
- native-iceberg-create — AWS Glue tarefa que cria uma tabela Iceberg no Catálogo de Dados
- native-hudi-create — AWS Glue tarefa que cria uma tabela Hudi no Catálogo de Dados
- native-delta-create — AWS Glue tarefa que cria uma tabela Delta no catálogo de dados
- LF-OTF-GlueServiceRole — Função do IAM para a qual você passa AWS Glue para executar os trabalhos. Essa perfil tem as políticas necessárias anexadas para acessar os recursos, como o catálogo de dados, o bucket do Amazon S3, etc.
- LF-OTF-RegisterRole — Função do IAM para registrar a localização do Amazon S3 no Lake Formation. Esse perfil tem o `LF-Data-Lake-Storage-Policy` anexado ao perfil.
- lf-consumer-analystuser — Usuário do IAM para consultar os dados usando o Athena

- lf-consumer-analystuser-credentials — Senha do usuário analista de dados armazenada em AWS Secrets Manager

Depois que as criações da pilha forem concluídas, navegue até a guia de saída e anote os valores para:

- AthenaQueryResultLocation — Localização do Amazon S3 para saída de consulta do Athena
- BusinessAnalystUserCredentials — Senha para o usuário analista de dados

  Como recuperar o valor da senha:

  1. Selecione o valor `lf-consumer-analystuser-credentials` navegando até o console do Secrets Manager.
  2. Na seção Valor secreto, selecione Recuperar valor secreto.
  3. Anote o valor secreto da senha.

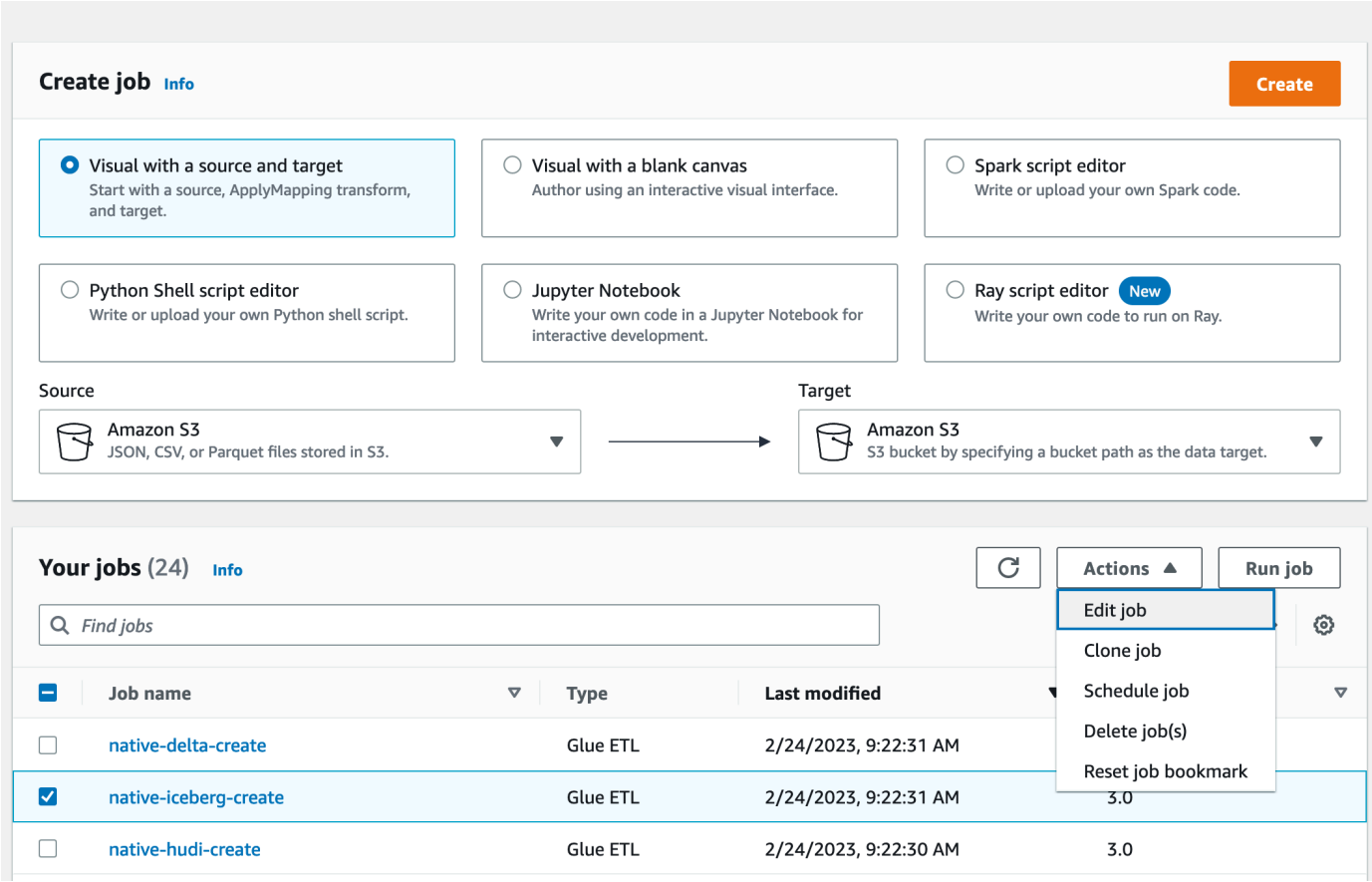
Etapa 2: Configurar permissões para uma tabela do Iceberg

Nesta seção, você aprenderá como criar uma tabela Iceberg no AWS Glue Data Catalog, configurar permissões de dados e consultar dados usando o Amazon Athena. AWS Lake Formation

Como criar uma tabela no Iceberg

Nesta etapa, você executará um AWS Glue trabalho que cria uma tabela transacional Iceberg no Catálogo de Dados.

1. Abra o AWS Glue console https://console.aws.amazon.com/glue/na região Leste dos EUA (Norte da Virgínia) como usuário administrador do data lake.
2. Selecione trabalhos no painel de navegação esquerdo.
3. Selecione `native-iceberg-create`.



4. Em Ações, selecione Editar trabalho.
5. Em Detalhes do trabalho, expanda Propriedades avançadas e marque a caixa ao lado de Usar AWS Glue Data Catalog como metastore do Hive para adicionar os metadados da tabela no. AWS Glue Data Catalog Isso especifica AWS Glue Data Catalog como metastore os recursos do Catálogo de Dados usados no trabalho e permite que as permissões do Lake Formation sejam aplicadas posteriormente nos recursos do catálogo.
6. Escolha Salvar.
7. Escolha Executar. Você pode exibir o status da tarefa durante sua execução.

   Para obter mais informações sobre AWS Glue trabalhos, consulte Como trabalhar com trabalhos no AWS Glue console no Guia do AWS Glue desenvolvedor.

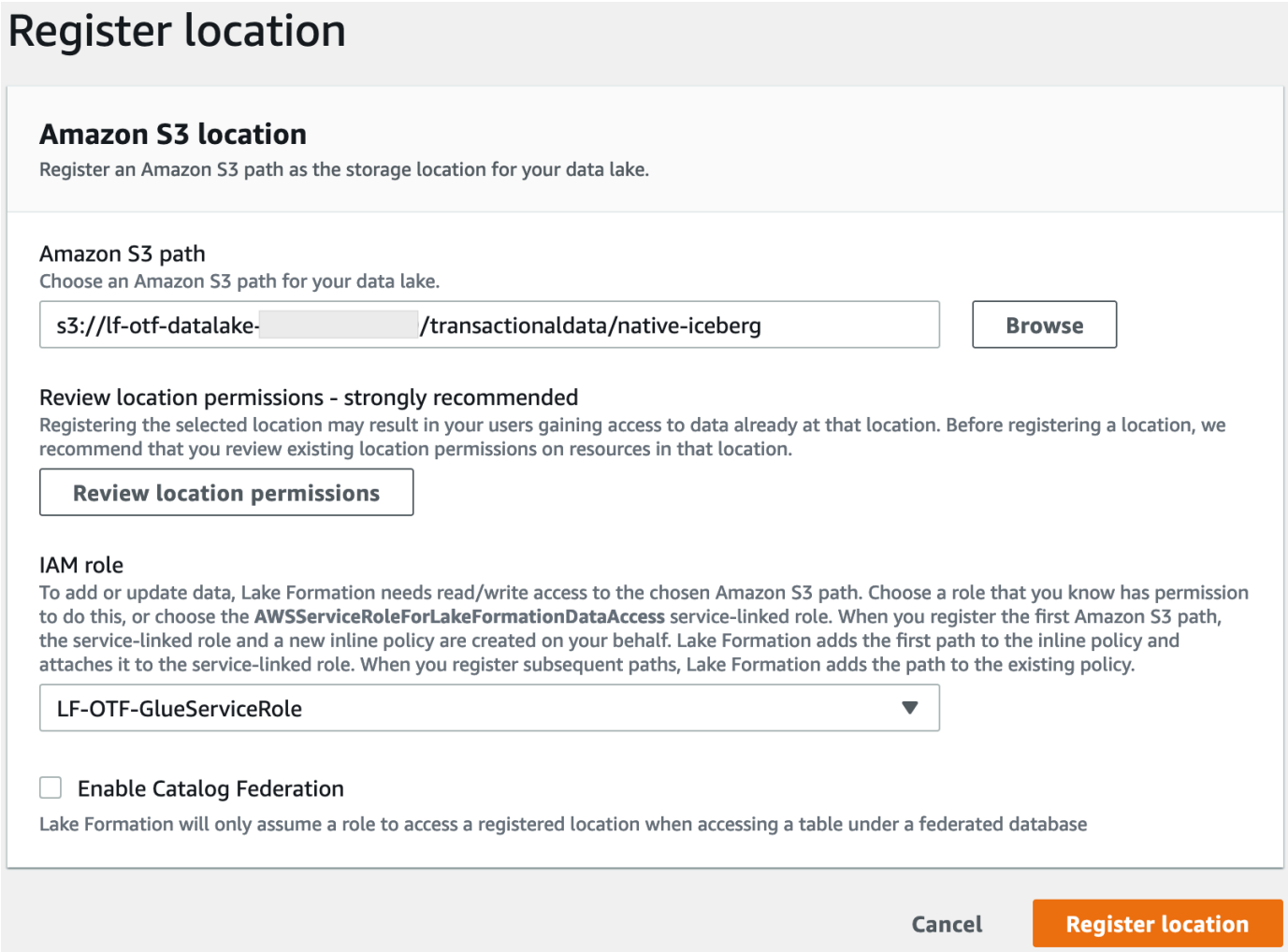
   Esse trabalho cria uma tabela no Iceberg nomeada como `product` banco de dados `lficebergdb`. Verifique a tabela de produtos no console do Lake Formation.

Como registrar o local dos dados com Lake Formation

Em seguida, registre o caminho do Amazon S3 como o local do seu data lake.

1. Abra o console do Lake Formation https://console.aws.amazon.com/lakeformation/como usuário administrador do data lake.
2. No painel de navegação, em Registrar e ingerir, selecione Local dos dados.
3. Selecione Registrar local.
4. Na página Registrar local, digite o seguinte:

   - Caminho do Amazon S3 — selecione Navegar e selecione `lf-otf-datalake-123456789012`. Escolha a seta de expansão ao lado da localização raiz do Amazon S3 para navegar até a `s3/buckets/lf-otf-datalake-123456789012/transactionaldata/native-iceberg` localização.
   - Perfil do IAM — selecione `LF-OTF-RegisterRole` como perfil do IAM.
   - Selecione Registrar local.

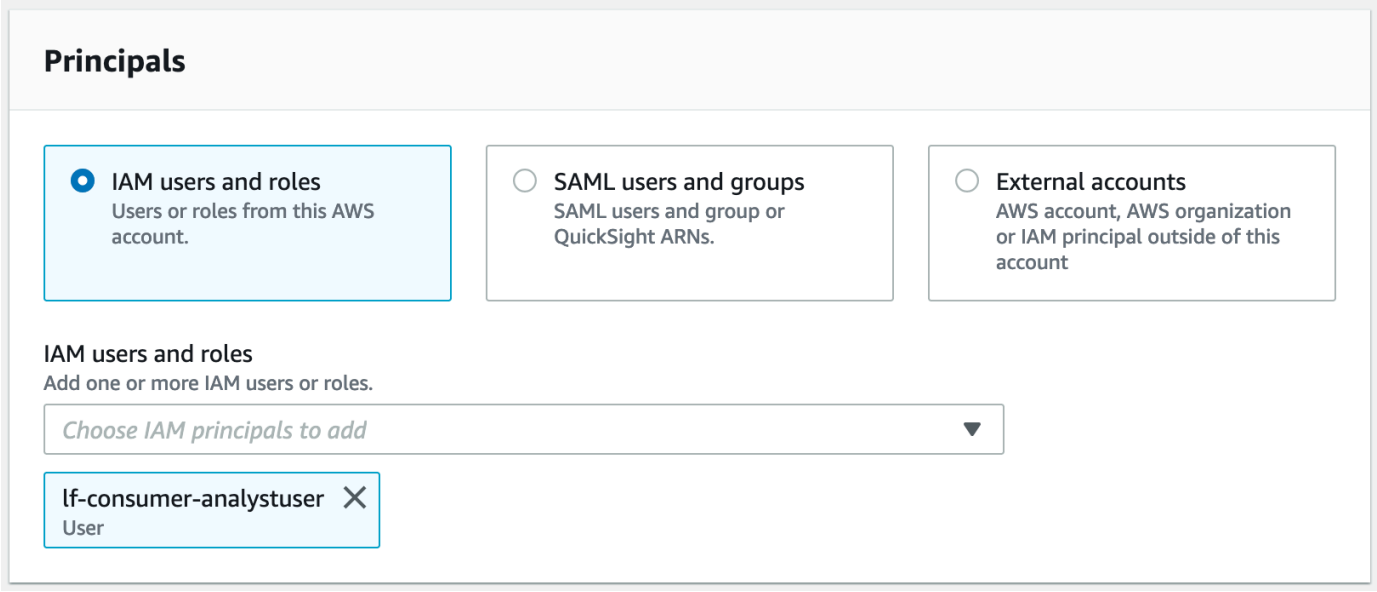


Para obter mais informações sobre como registrar um local de dados no Lake Formation, consulte Adicionar uma localização do Amazon S3 ao seu data lake.

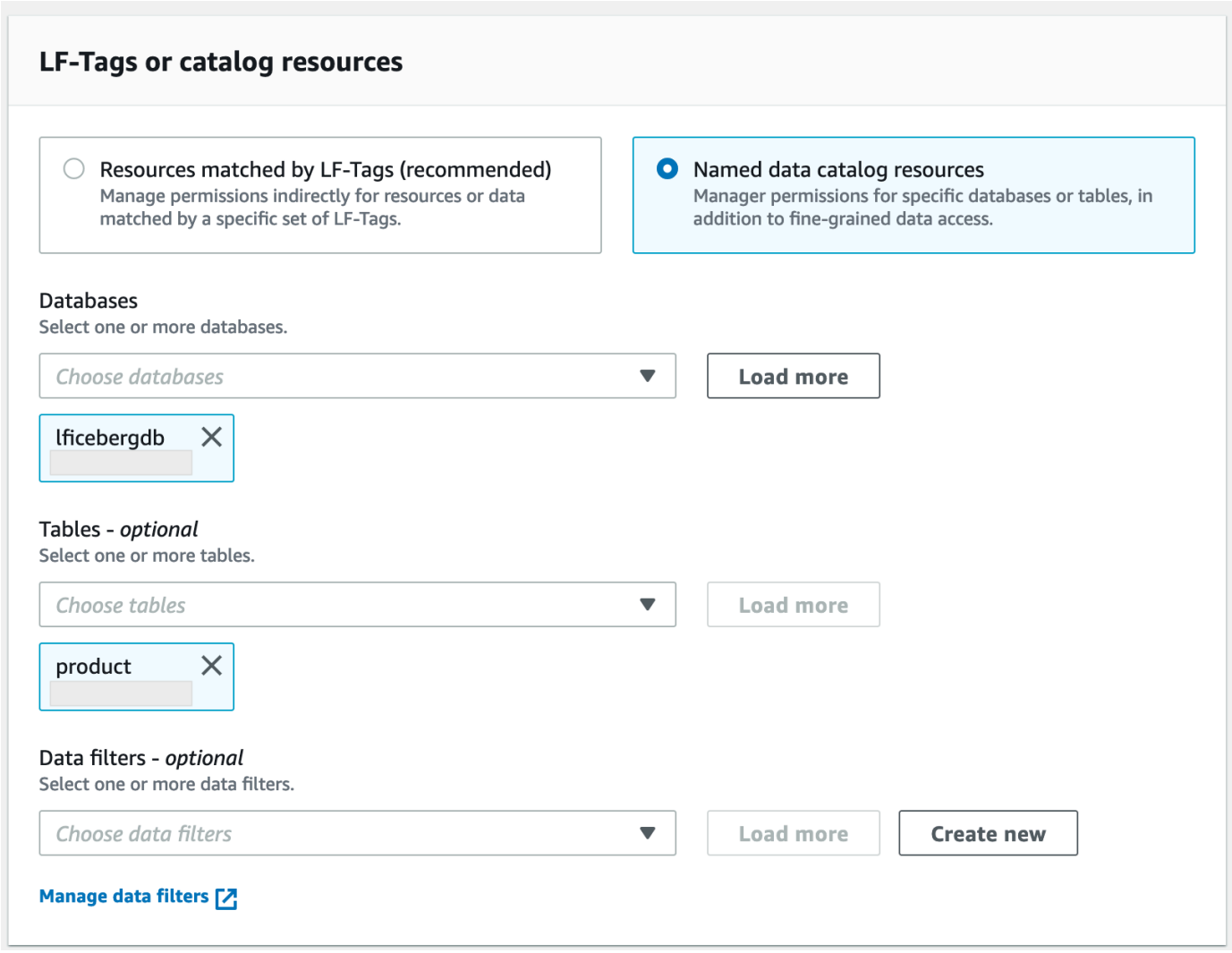
Como conceder permissões do Lake Formation na tabela do Iceberg

Nesta etapa, concederemos permissões de data lake ao usuário do analista de negócios.

1. Em Permissões do Data Lake, selecione Conceder.
2. Na tela Conceder permissões de dados, selecione Usuários e perfis do IAM.
3. Selecione `lf-consumer-analystuser` no menu suspenso.



4. Selecione Recurso nomeado de catálogo de dados.
5. Para o Banco de dados, selecione `lficebergdb`.
6. Para Tabelas, selecione `product`.



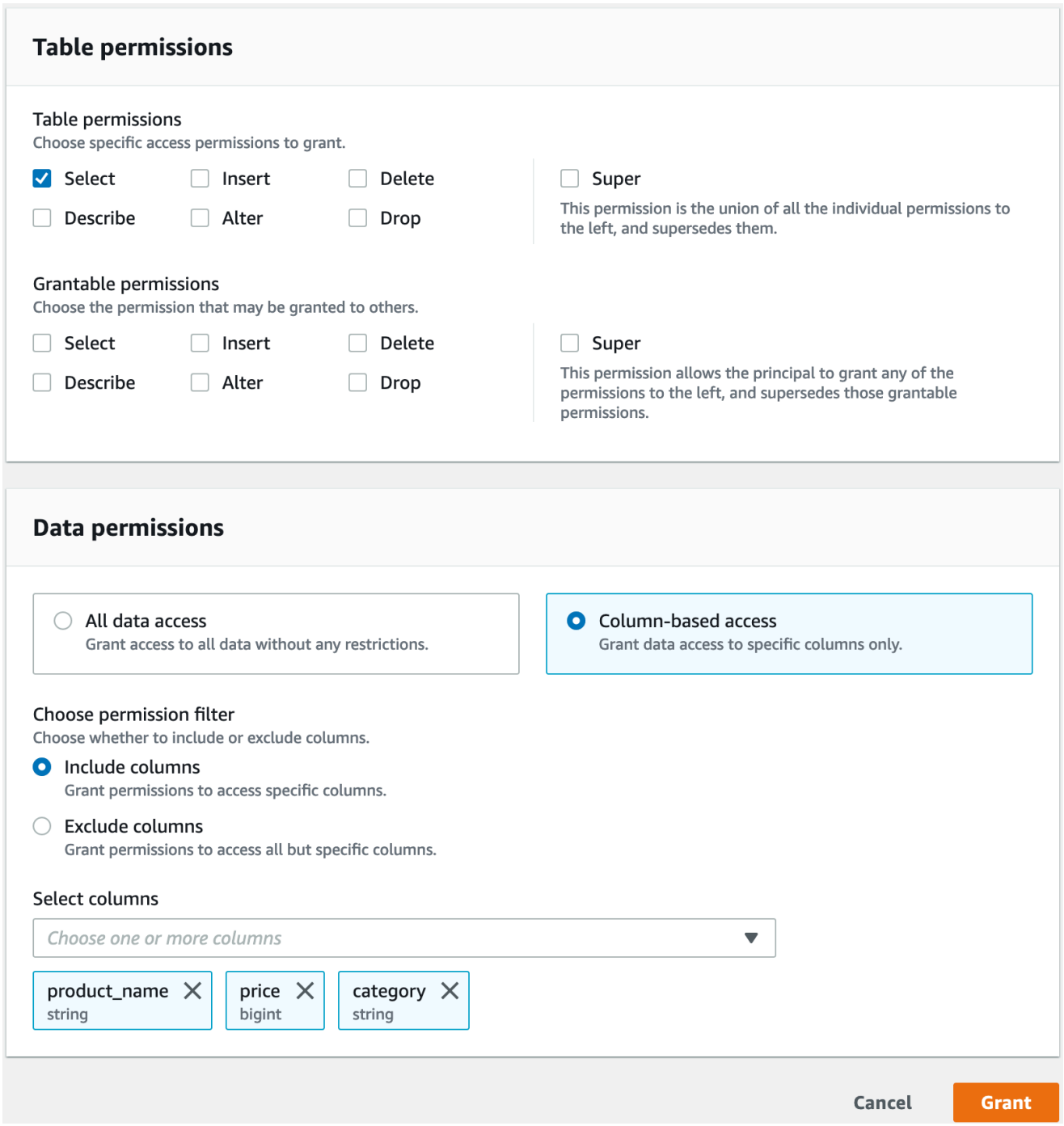
7. Em seguida, você pode conceder acesso baseado em colunas especificando colunas.

   a. Em Permissões de tabela, clique em Selecionar.

   b. Em Permissões de dados, escolha Column-based Acesso, escolha Incluir colunas.

   c. Selecione as colunas `product_name`, `price` e `category`.

   d. Selecione Conceder.



Para consultar a tabela do Iceberg usando o Athena

Agora você pode começar a consultar a tabela do Iceberg que você criou usando o Athena. Se for a primeira vez que você executa consultas no Athena, você precisará configurar o local do resultado da consulta. Para obter mais informações, acesse Especificação de um local de resultado de consulta.

1. Saia como usuário administrador do data lake e faça login como `lf-consumer-analystuser` na região Leste dos EUA (Norte da Virgínia) usando a senha anotada anteriormente na CloudFormation saída.
2. Abra o console do Athena em https://console.aws.amazon.com/athena/.
3. Selecione Configurações e selecione Gerenciar.
4. Na caixa Localização do resultado da consulta, insira o caminho para o bucket que você criou nas CloudFormation saídas. Copie o valor de **`AthenaQueryResultLocation`** (s3://lf-otf-tutorial - 123456789012/athena -results/) e escolha Salvar.
5. Execute a consulta a seguir para exibir 10 registros armazenados na tabela do Iceberg:

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

   Para obter mais informações sobre como consultar tabelas do Iceberg usando o Athena, consulte Como consultar tabelas do Iceberg no Guia do usuário do Amazon Athena.

Etapa 3: Configurar permissões para uma tabela do Hudi

Nesta seção, você aprenderá como criar uma tabela Hudi no AWS Glue Data Catalog, configurar permissões de dados e consultar dados usando o Amazon Athena. AWS Lake Formation

Como criar uma tabela Hudi

Nesta etapa, você executará um AWS Glue trabalho que cria uma tabela transacional Hudi no Catálogo de Dados.

1. Faça login no AWS Glue console https://console.aws.amazon.com/glue/na região Leste dos EUA (Norte da Virgínia)

   como o usuário administrador do data lake.
2. Selecione trabalhos no painel de navegação esquerdo.
3. Selecione `native-hudi-create`.
4. Em Ações, selecione Editar trabalho.
5. Em Detalhes do trabalho, expanda Propriedades avançadas e marque a caixa ao lado de Usar AWS Glue Data Catalog como metastore do Hive para adicionar os metadados da tabela no. AWS Glue Data Catalog Isso especifica AWS Glue Data Catalog como metastore os recursos do

Catálogo de Dados usados no trabalho e permite que as permissões do Lake Formation sejam aplicadas posteriormente nos recursos do catálogo.

6. Escolha Salvar.
7. Escolha Executar. Você pode exibir o status da tarefa durante sua execução.

   Para obter mais informações sobre AWS Glue trabalhos, consulte Como trabalhar com trabalhos no AWS Glue console no Guia do AWS Glue desenvolvedor.

   Esse trabalho cria uma tabela do Hudi (COW) no banco de dados: lfhudidb. Verifique a tabela no console `product` do Lake Formation.

Como registrar o local dos dados com Lake Formation

Em seguida, registre um caminho do Amazon S3 como o local raiz do data lake.

1. Faça login no console do Lake Formation https://console.aws.amazon.com/lakeformation/como usuário administrador do data lake.
2. No painel de navegação, em Registrar e ingerir, selecione Local dos dados.
3. Selecione Registrar local.
4. Na página Registrar local, digite o seguinte:

   - Caminho do Amazon S3 — selecione Navegar e selecione `lf-otf-datalake-123456789012`. Escolha a seta de expansão ao lado da localização raiz do Amazon S3 para navegar até a `s3/buckets/lf-otf-datalake-123456789012/transactionaldata/native-hudi` localização.
   - Perfil do IAM — selecione `LF-OTF-RegisterRole` como perfil do IAM.
   - Selecione Registrar local.

Como conceder permissões de data lake na tabela do Hudi

Nesta etapa, concederemos permissões de data lake ao usuário do analista de negócios.

1. Em Permissões do Data Lake, selecione Conceder.
2. Na tela Conceder permissões de dados, selecione Usuários e perfis do IAM.
3. `lf-consumer-analystuser` do menu suspenso.
4. Selecione Recurso nomeado de catálogo de dados.

5. Para o Banco de dados, selecione `lfhudidb`.
6. Para Tabelas, selecione `product`.
7. Em seguida, você pode conceder acesso baseado em colunas especificando colunas.

   a. Em Permissões de tabela, clique em Selecionar.
   b. Em Permissões de dados, escolha Column-based Acesso, escolha Incluir colunas.
   c. Selecione as colunas `product_name`, `price` e `category`.
   d. Selecione Conceder.

Como consultar a tabela do Hudi usando o Athena

Agora comece a consultar a tabela do Hudi que você criou usando o Athena. Se for a primeira vez que você executa consultas no Athena, você precisará configurar o local do resultado da consulta. Para obter mais informações, acesse Especificação de um local de resultado de consulta.

1. Saia como usuário administrador do data lake e faça login como `lf-consumer-analystuser` na região Leste dos EUA (Norte da Virgínia) usando a senha anotada anteriormente na CloudFormation saída.
2. Abra o console do Athena em https://console.aws.amazon.com/athena/.
3. Selecione Configurações e selecione Gerenciar.
4. Na caixa Localização do resultado da consulta, insira o caminho para o bucket que você criou nas CloudFormation saídas. Copie o valor de **`AthenaQueryResultLocation`** (s3://lf-otf-tutorial - 123456789012/athena -results/) e salve.
5. Execute a consulta a seguir para exibir 10 registros armazenados na tabela do Hudi:

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

   Para obter mais informações sobre como consultar tabelas do Hudi usando o Athena, consulte Como consultar tabelas do Hudi no Guia do usuário do Amazon Athena.

Etapa 4: Configurar permissões para uma tabela do Delta Lake

Nesta seção, você aprenderá como criar uma tabela Delta Lake com arquivo de manifesto de link simbólico no AWS Glue Data Catalog, configurar permissões de dados AWS Lake Formation e consultar dados usando o Amazon Athena.

Como criar uma tabela do Delta Lake

Nesta etapa, você executará um AWS Glue trabalho que cria uma tabela transacional do Delta Lake no Catálogo de Dados.

1. Faça login no AWS Glue console https://console.aws.amazon.com/glue/na região Leste dos EUA (Norte da Virgínia)

   como o usuário administrador do data lake.
2. Selecione trabalhos no painel de navegação esquerdo.
3. Selecione `native-delta-create`.
4. Em Ações, selecione Editar trabalho.
5. Em Detalhes do trabalho, expanda Propriedades avançadas e marque a caixa ao lado de Usar AWS Glue Data Catalog como metastore do Hive para adicionar os metadados da tabela no. AWS Glue Data Catalog Isso especifica AWS Glue Data Catalog como metastore os recursos do Catálogo de Dados usados no trabalho e permite que as permissões do Lake Formation sejam aplicadas posteriormente nos recursos do catálogo.
6. Escolha Salvar.
7. Selecione Executar em Ações.

   Esse trabalho cria uma tabela no Delta Lake nomeada como `product` banco de dados `lfdeltadb`. Verifique a tabela no console `product` do Lake Formation.

Como registrar o local dos dados com Lake Formation

Em seguida, registre o caminho do Amazon S3 como o local raiz do data lake.

1. Abra o console do Lake Formation no https://console.aws.amazon.com/lakeformation/usuário administrador do data lake.
2. No painel de navegação, em Registrar e ingerir, selecione Local dos dados.
3. Selecione Registrar local.
4. Na página Registrar local, digite o seguinte:

   - Caminho do Amazon S3 — selecione Navegar e selecione `lf-otf-datalake-123456789012`. Escolha a seta de expansão ao lado da localização raiz do Amazon S3 para navegar até a `s3/buckets/lf-otf-datalake-123456789012/transactionaldata/native-delta` localização.

- Perfil do IAM — selecione `LF-OTF-RegisterRole` como perfil do IAM.
- Selecione Registrar local.

Como conceder permissões de data lake na tabela do Delta Lake

Nesta etapa, concederemos permissões de data lake ao usuário do analista de negócios.

1. Em Permissões do Data Lake, selecione Conceder.
2. Na tela Conceder permissões de dados, selecione Usuários e perfis do IAM.
3. `lf-consumer-analystuser` do menu suspenso.
4. Selecione Recurso nomeado de catálogo de dados.
5. Para o Banco de dados, selecione `lfdeltadb`.
6. Para Tabelas, selecione `product`.
7. Em seguida, você pode conceder acesso baseado em colunas especificando colunas.

    a. Em Permissões de tabela, clique em Selecionar.

    b. Em Permissões de dados, escolha Column-based Acesso, escolha Incluir colunas.

    c. Selecione as colunas `product_name`, `price` e `category`.

    d. Selecione Conceder.

Como consultar a tabela do Delta Lake usando o Athena

Agora comece a consultar a tabela do Delta Lake que você criou usando o Athena. Se for a primeira vez que você executa consultas no Athena, você precisará configurar o local do resultado da consulta. Para obter mais informações, acesse Especificação de um local de resultado de consulta.

1. Saia como usuário administrador do data lake e faça login como `BusinessAnalystUser` na região Leste dos EUA (Norte da Virgínia) usando a senha anotada anteriormente na CloudFormation saída.
2. Abra o console do Athena em https://console.aws.amazon.com/athena/.
3. Selecione Configurações e selecione Gerenciar.
4. Na caixa Localização do resultado da consulta, insira o caminho para o bucket que você criou nas CloudFormation saídas. Copie o valor de **`AthenaQueryResultLocation`** (s3://lf-otf-tutorial - 123456789012/athena -results/) e salve.
5. Execute a consulta a seguir para exibir 10 registros armazenados na tabela do Delta Lake:

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

Para obter mais informações sobre como consultar tabelas do Delta Lake usando o Athena, consulte Como consultar tabelas do Delta Lake no Guia do usuário do Amazon Athena.

Etapa 5: limpar AWS recursos

Como limpar recursos

Para evitar cobranças indesejadas Conta da AWS, exclua AWS os recursos que você usou neste tutorial.

1. Faça login no CloudFormation console https://console.aws.amazon.com/cloudformationcomo administrador do IAM.
2. Exclua a pilha do Cloud Formation. As tabelas que você criou são excluídas automaticamente com a pilha.

Como gerenciar um data lake usando o controle de acesso baseado em tags do Lake Formation

Milhares de clientes estão criando lagos de dados em escala de petabytes. AWS Muitos desses clientes usam AWS Lake Formation para criar e compartilhar facilmente seus lagos de dados em toda a organização. À medida que o número de tabelas e usuários aumenta, administradores de dados procuram maneiras de gerenciar facilmente as permissões em data lakes em grande escala. O controle de acesso baseado em Lake Formation tags (LF-TBAC) resolve essa questão, permitindo que administradores de dados criem tags do LF (com base em sua classificação e ontologia de dados) que podem ser anexadas aos recursos.

O LF-TBAC é uma estratégia de autorização que define permissões com base em atributos. No Lake Formation, esses atributos são chamados de tags do LF. Você pode anexar tags do LF aos recursos do catálogo de dados e às entidades principais do Lake Formation. Administradores do Data Lake podem atribuir e revogar permissões nos recursos do Lake Formation usando tags do LF. Para obter mais informações, consulte Controle de acesso baseado em tags do Lake Formation.

Este tutorial demonstra como criar uma política de controle de acesso baseada em tags do Lake Formation usando um conjunto de dados AWS público. Além disso, mostra como consultar tabelas,

bancos de dados e colunas que têm políticas de acesso baseadas em tags do Lake Formation associadas a eles.

Você pode usar o LF-TBAC para os seguintes casos:

- Você tem um grande número de tabelas e entidades principais às quais o administrador do data lake precisa conceder acesso
- Você deseja classificar seus dados com base em uma ontologia e conceder permissões com base na classificação
- O administrador do data lake deseja atribuir permissões dinamicamente, com acoplamento fraco

A seguir estão as etapas de alto nível para configurar as permissões usando o LF-TBAC:

1. O administrador de dados define a ontologia da tag com duas tags do LF: `Confidential` e `Sensitive`. Os dados com `Confidential=True` têm controles de acesso mais rígidos. Os dados com `Sensitive=True` requerem uma análise específica do analista.
2. O administrador de dados atribui diferentes níveis de permissão ao engenheiro de dados para criar tabelas com diferentes tags do LF.
3. O engenheiro de dados cria dois bancos de dados: `tag_database` e `col_tag_database`. Todas as tabelas em `tag_database` são configuradas com `Confidential=True`. Todas as tabelas do `col_tag_database` são configuradas com `Confidential=False`. Algumas colunas da tabela `col_tag_database` estão marcadas com `Sensitive=True` para necessidades específicas de análise.
4. O engenheiro de dados concede permissão de leitura ao analista para tabelas com condições de expressão específicas `Confidential=True` e `Confidential=False`, `Sensitive=True`.
5. Com essa configuração, o analista de dados pode se concentrar em realizar análises com os dados certos.

Tópicos

- Público-alvo
- Pré-requisitos
- Etapa 1: Provisionar os recursos
- Etapa 2: registrar sua localização de dados, criar uma ontologia da tag do LF e conceder permissões
- Etapa 3: Criar bancos de dados do Lake Formation

- Etapa 4: Conceder permissões de dados
- Etapa 5: Executar uma consulta no Amazon Athena para verificar as permissões
- Etapa 6: limpar AWS os recursos

Público-alvo

Este tutorial é destinado a administradores de dados, engenheiros de dados e analistas de dados. Quando se trata de gerenciar AWS Glue Data Catalog e administrar permissões no Lake Formation, os administradores de dados nas contas produtoras têm propriedade funcional com base nas funções que suportam e podem conceder acesso a vários consumidores, organizações externas e contas.

A tabela a seguir lista os perfis usados neste tutorial:

Perfil	Description
Administrador de dados (administrador)	O usuário `lf-data-steward` tem o seguinte acesso: • Acesso de leitura a todos os recursos do catálogo de dados • Pode criar tags do LF e associar-se ao perfil de engenheiro de dados para conceder permissão a outras entidades principais
Engenheiro de dados	O usuário `lf-data-engineer` tem o seguinte acesso: • Acesso completo de leitura, gravação e atualização a todos os recursos do catálogo de dados • Permissões de local de dados no data lake • Pode associar tags do LF e associar-se ao catálogo de dados • Pode anexar tags do LF aos recursos, fornecendo acesso às entidades principais

Perfil	Description
	com base em quaisquer políticas criadas por administradores de dados
Analista de dados	O usuário `lf-data-analyst` tem o seguinte acesso: • Acesso refinado aos recursos compartilhados pelas políticas de acesso baseadas em tags do Lake Formation

Pré-requisitos

Antes de começar este tutorial, você deve ter um Conta da AWS que possa ser usado para entrar como usuário administrativo com as permissões corretas. Para obter mais informações, consulte Inicial completa AWS tarefas de configuração.

O tutorial pressupõe que você esteja familiarizado com o IAM. Para obter informações sobre o IAM, consulte o Guia do usuário do IAM.

Etapa 1: Provisionar os recursos

Este tutorial inclui um AWS CloudFormation modelo para uma configuração rápida. É possível revisá-lo e personalizá-lo para atender às suas necessidades. O modelo cria três funções diferentes (listadas emPúblico-alvo) para realizar esse exercício e copia o nyc-taxi-data conjunto de dados para seu bucket local do Amazon S3.

- Um bucket Amazon S3
- Configurações apropriadas do Lake Formation
- Recursos apropriados do Amazon EC2
- Três perfis do IAM com credenciais

Criar seus recursos

1. Faça login no AWS CloudFormation console em https://console.aws.amazon.com/cloudformation na região Leste dos EUA (Norte da Virgínia).

2. Selecione Iniciar Pilha.
3. Escolha Próximo.
4. Na seção Configuração do usuário, digite a senha para três perfis: `DataStewardUserPassword`, `DataEngineerUserPassword` e `DataAnalystUserPassword`.
5. Analise os detalhes na página final e selecione Eu reconheço que isso AWS CloudFormation pode criar recursos do IAM.
6. Escolha Criar.

   A criação da pilha pode levar até cinco minutos.

> **Note**
>
> Depois de concluir o tutorial, talvez você queira excluir a pilha CloudFormation para evitar que continuem incorrendo em cobranças. Verifique se os recursos foram excluídos com sucesso no status de eventos da pilha.

Etapa 2: registrar sua localização de dados, criar uma ontologia da tag do LF e conceder permissões

Nesta etapa, o usuário administrador de dados define a ontologia de tags com duas tags do LF (`Confidential` e `Sensitive`) e possibilita que entidades principais específicas do IAM anexem tags do LF recém-criadas aos recursos.

Registre uma localização de dados e defina a ontologia da tag do LF

1. Execute a primeira etapa como usuário administrador de dados (`lf-data-steward`) para verificar os dados no Amazon S3 e no catálogo de dados no Lake Formation.

   a. Faça login no console do Lake Formation da https://console.aws.amazon.com/lakeformation/mesma forma que `lf-data-steward` com a senha usada ao implantar a CloudFormation pilha.

   b. No painel de navegação, em Permissões, selecione Perfis e tarefas administrativas.

   c. Na seção Administradores do data lake, selecione Adicionar.

d. Na página Adicionar administrador, em Usuários e perfis do IAM, escolha o usuário `lf-data-steward`.

e. selecione Salvar para adicionar `lf-data-steward` como administrador do Lake Formation.

2. Em seguida, atualize as configurações do catálogo de dados para usar a permissão do Lake Formation para controlar os recursos do catálogo em vez do controle de acesso baseado no IAM.

a. No painel de navegação, em Administração, selecione Configurações do catálogo de dados.

b. Desmarque Usar somente o controle de acesso do IAM para novos bancos de dados.

c. Desmarque Usar somente o controle de acesso do IAM para novas tabelas em novos bancos de dados.

d. Clique em Salvar.

3. Em seguida, registre o local dos dados para o data lake.

a. No painel de navegação, em Administração em Locais de data lake.

b. Selecione Registrar local.

c. Na página Registrar localização, em Caminho do Amazon S3, insira `s3://lf-tagbased-demo-`*`Account-ID`*.

d. Para o Perfil do IAM, deixe o valor padrão `AWSServiceRoleForLakeFormationDataAccess` como está.

e. Escolha Lake Formation como o modo de permissão.

f. Selecione Registrar local.

4. Em seguida, crie a ontologia definindo uma tag do LF.

a. No painel de navegação, em Permissões, selecione Tags do LF e permissões.

b. Selecione Adicionar tag do LF.

c. Em Chave, digite `Confidential`.

d. Para Valores, adicione `True` e `False`.

e. Selecione Adicionar tag do LF.

f. Repita as etapas para criar a tag do LF `Sensitive` com o valor `True`.

Você criou todas as tags do LF necessárias para este exercício.

Conceda permissões a usuários do IAM

1. Em seguida, forneça às entidades principais específicas do IAM a capacidade de anexar tags do LF recém-criadas aos recursos.

   a. No painel de navegação, em Permissões, selecione Tags do LF e permissões.

   b. Na seção Permissões da tag do LF, escolha Conceder permissões.

   c. Em Tipo de permissão, escolha Permissões do par de chave-valor da tag do LF.

   d. Selecione Usuários e perfis do IAM.

   e. Para Usuários e perfis do IAM, pesquise e selecione o perfil `lf-data-engineer`.

   f. Na seção Tags do LF, adicione a chave `Confidential` com valores `True` e `False`, e a `key Sensitive` com valor `True`.

   g. Em Permissões, selecione Descrever e Associar para Permissões e Permissões concedíveis.

   h. Selecione Conceder.

2. Em seguida, conceda permissões `lf-data-engineer` para criar bancos de dados em nosso catálogo de dados e no bucket subjacente do Amazon S3 criado por. AWS CloudFormation

   a. No painel de navegação, em Administração, selecione Perfis e tarefas administrativas.

   b. Na seção Criadores de banco de dados, selecione Conceder.

   c. Para Usuários e perfis do IAM, selecione o perfil `lf-data-engineer`.

   d. Para Permissões de catálogo, selecione Criar banco de dados.

   e. Selecione Conceder.

3. Em seguida, conceda permissões no bucket do Amazon S3 do (`s3://lf-tagbased-demo-`*`Account-ID`*`)` ao usuário `lf-data-engineer`.

   a. No painel de navegação, em Permissões, selecione Locais de dados.

   b. Selecione Conceder.

   c. Selecione Minha conta.

   d. Para Usuários e perfis do IAM, selecione o perfil `lf-data-engineer`.

   e. Para locais de armazenamento, insira o bucket do Amazon S3 criado pelo CloudFormation modelo. (`s3://lf-tagbased-demo-`*`Account-ID`*`)`

   f. Selecione Conceder.

4. Em seguida, conceda permissões concedíveis `lf-data-engineer` aos recursos associados à expressão `Confidential=True` da tag do LF.

   a. No painel de navegação, em Permissões, escolha Permissões do data lake.

   b. Selecione Conceder.

   c. Selecione Usuários e perfis do IAM.

   d. Selecione o perfil de `lf-data-engineer`.

   e. Na seção Tag do LF ou recursos do catálogo, selecione Recursos correspondentes às tags do LF.

   f. Escolha Adicionar par de chave-valor da tag do LF.

   g. Adicione a chave `Confidential` com os valores `True`.

   h. Na seção Permissões do banco de dados, selecione Descrever em Permissões de banco de dados e Permissões concedíveis.

   i. Na seção Permissões de tabela, selecione Descrever, Selecionar e Alterar para as Permissões de tabela e as Permissões concedíveis.

   j. Selecione Conceder.

5. Em seguida, conceda as permissões concedíveis `lf-data-engineer` aos recursos associados à expressão `Confidential=False` da tag do LF.

   a. No painel de navegação, em Permissões, escolha Permissões do data lake.

   b. Selecione Conceder.

   c. Selecione Usuários e perfis do IAM.

   d. Selecione o perfil de `lf-data-engineer`.

   e. Selecione Recursos correspondentes às tags do LF.

   f. Selecione Adicionar tag do LF.

   g. Adicione a chave `Confidential` com o valor `False`.

   h. Na seção Permissões do banco de dados, selecione Descrever em Permissões de banco de dados e Permissões concedíveis.

   i. Na seção Permissões de tabela e coluna, não selecione nada.

   j. Selecione Conceder.

6. Em seguida, damos permissões concedíveis `lf-data-engineer` aos recursos associados aos pares de chave-valor `Confidential=False` e `Sensitive=True` da tag do LF.

a. No painel de navegação, em Permissões, escolha Permissões de dados.
b. Selecione Conceder.
c. Selecione Usuários e perfis do IAM.
d. Selecione o perfil de `lf-data-engineer`.
e. Na seção Tags do LF ou recursos do catálogo, selecione Recursos correspondentes às tags do LF.
f. Selecione Adicionar tag do LF.
g. Adicione a chave `Confidential` com o valor `False`.
h. Escolha Adicionar par de chave-valor da tag do LF.
i. Adicione a chave `Sensitive` com o valor `True`.
j. Na seção Permissões do banco de dados, selecione Descrever em Permissões de banco de dados e Permissões concedíveis.
k. Na seção Permissões de tabela, selecione Descrever, Selecionar e Alterar para as Permissões de tabela e as Permissões concedíveis.
l. Selecione Conceder.

Etapa 3: Criar bancos de dados do Lake Formation

Nesta etapa, você cria dois bancos de dados e anexa tags do LF aos bancos de dados e às colunas específicas para fins de teste.

Crie seus bancos de dados e sua tabela para acesso em nível de banco de dados

1. Primeiro, crie o banco de dados `tag_database` e a tabela `source_data` e anexe as tags do LF apropriadas.

   a. No console do Lake Formation (https://console.aws.amazon.com/lakeformation/), em Catálogo de dados, escolha Bancos de dados.
   b. Selecione Criar banco de dados.
   c. Em Nome, digite `tag_database`.
   d. Em Localização, insira a localização do Amazon S3 criada pelo CloudFormation modelo. (`s3://lf-tagbased-demo-`*`Account-ID`*`/tag_database/`)
   e. Desmarque Usar somente controle de acesso do IAM para novas tabelas nesse banco de dados.

f. Selecione Criar banco de dados.

2. Em seguida, crie uma nova tabela dentro dela com `tag_database`.

   a. Na página Bancos de dados, selecione o banco de dados `tag_database`.

   b. Selecione Exibir tabelas e clique em Criar tabela.

   c. Em Nome, digite `source_data`.

   d. Em Banco de dados, selecione o banco de dados `tag_database`.

   e. Em Formato de tabela, escolha AWS Glue Tabela padrão.

   f. Em Dados localizados em, selecione Caminho especificado em minha conta.

   g. Em Incluir caminho, insira o caminho a ser `tag_database` criado pelo CloudFormation modelo(`s3://lf-tagbased-demo`*`Account-ID`*`/tag_database/`).

   h. Em Formato de dados, selecione CSV.

   i. Em Esquema de upload, digite a seguinte matriz JSON da estrutura da coluna para criar um esquema:

```
[
              {
                   "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. Selecione Carregar. Após fazer o upload do esquema, o esquema da tabela deve ter a aparência da seguinte captura de tela:

#	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. Selecione Enviar.

3. Em seguida, anexe as tags do LF no nível do banco de dados.

   a. Na página Bancos de dados, localize e selecione `tag_database`.
   b. No menu Ações, selecione Editar tags do LF.
   c. Escolha Atribuir nova tag do LF.
   d. Em Chaves atribuídas, selecione a tag do LF `Confidential` que você criou anteriormente.
   e. Em Valores, selecione `True`.
   f. Selecione Salvar.

   Isso conclui a atribuição de tags do LF ao banco de dados tag_database.

Crie seu banco de dados e tabela para acesso em nível de coluna

Repita as etapas a seguir para criar o banco de dados `col_tag_database` e a tabela `source_data_col_lvl` e anexar tags do LF no nível da coluna.

1. Na página Bancos de dados, selecione Criar banco de dados.
2. Em Nome, digite `col_tag_database`.
3. Em Localização, insira a localização do Amazon S3 criada pelo CloudFormation modelo. (`s3://lf-tagbased-demo-`*`Account-ID`*`/col_tag_database/`)
4. Desmarque Usar somente controle de acesso do IAM para novas tabelas nesse banco de dados.
5. Selecione Criar banco de dados.
6. Na página Bancos de dados, selecione seu novo banco de dados (`col_tag_database`).
7. Selecione Visualizar tabelas e clique em Criar tabela.
8. Em Nome, digite `source_data_col_lvl`.
9. Em Banco de dados, selecione seu novo banco de dados (`col_tag_database`).
10. Em Formato de tabela, escolha AWS Glue Tabela padrão.
11. Em Dados localizados em, selecione Caminho especificado em minha conta.
12. Digite o caminho do Amazon S3 para `col_tag_database` (`s3://lf-tagbased-demo-`*`Account-ID`*`/col_tag_database/`).
13. Em Formato de dados, selecione `CSV`.

14. Em `Upload schema`, digite o seguinte esquema JSON:

```
[
                {
                     "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. Selecione `Upload`. Após fazer o upload do esquema, o esquema da tabela deve ter a aparência da seguinte captura de tela.

#	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. Selecione Enviar para concluir a criação da tabela.
17. Agora associe a tag do LF `Sensitive=True` às colunas `vendorid` e `fare_amount`.
    a. Na página Tabelas, selecione a tabela que você criou (`source_data_col_lvl`).
    b. No menu Ações, selecione Esquema.
    c. Selecione a coluna `vendorid` e escolha Editar tags do LF.
    d. Em Chaves atribuídas, selecione Sensível.
    e. Em Valores, selecione True.
    f. Escolha Salvar.
18. Em seguida, associe a tag do LF `Confidential=False` ao `col_tag_database`. Isso é necessário `lf-data-analyst` para poder descrever o banco de dados `col_tag_database` quando conectado. Amazon Athena
    a. Na página Bancos de dados, localize e selecione `col_tag_database`.
    b. No menu Ações, selecione Editar tags do LF.
    c. Escolha Atribuir nova tag do LF.
    d. Em Chaves atribuídas, selecione a tag do LF `Confidential` que você criou anteriormente.
    e. Em Valores, selecione `False`.
    f. Escolha Salvar.

Etapa 4: Conceder permissões de dados

Conceda permissões aos analistas de dados para o consumo dos bancos de dados `tag_database` e da tabela `col_tag_database` usando as tags do LF `Confidential` e `Sensitive`.

1. Siga estas etapas para conceder permissões ao usuário `lf-data-analyst` nos objetos associados à tag do LF `Confidential=True` (Database:tag_database) para ter a permissão `Describe` no banco de dados e `Select` nas tabelas.
    a. Faça login no console do Lake Formation em https://console.aws.amazon.com/lakeformation/como`lf-data-engineer`.
    b. Em Permissões, escolha Permissões do data lake.
    c. Selecione Conceder.
    d. Em Entidades principais, selecione Usuários e perfis do IAM.

e. Para Usuários e perfis do IAM, selecione `lf-data-analyst`.

f. Em Tags do LF ou recursos do catálogo, selecione Recursos correspondentes às tags do LF.

g. Selecione Adicionar tag do LF.

h. Para Chave, selecione `Confidential`.

i. Em Valores, selecione `True`.

j. Para Permissões de banco de dados, selecione `Describe`.

k. Para Permissões de tabela, clique em Selecionar e Descrever.

l. Selecione Conceder.

2. Em seguida, repita as etapas para conceder permissões aos analistas de dados referentes à expressão da tag do LF para `Confidential=False`. Essa tag do LF é usada para descrever o `col_tag_database` e a tabela `source_data_col_lvl` quando conectada como `lf-data-analyst` no Amazon Athena.

   a. Faça login no console do Lake Formation em https://console.aws.amazon.com/lakeformation/como`lf-data-engineer`.

   b. Na página Bancos de dados, selecione o banco de dados `col_tag_database`.

   c. Selecione Ações e Concessão.

   d. Em Entidades principais, selecione Usuários e perfis do IAM.

   e. Para Usuários e perfis do IAM, selecione `lf-data-analyst`.

   f. Selecione Recursos correspondentes às tags do LF.

   g. Selecione Adicionar tag do LF.

   h. Para Chave, selecione `Confidential`.

   i. Para Valores, selecione `False`.

   j. Para Permissões de banco de dados, selecione `Describe`.

   k. Para Permissões de tabela, não selecione nada.

   l. Selecione Conceder.

3. Em seguida, repita as etapas para conceder permissões aos analistas de dados referentes à expressão da tag do LF para `Confidential=False` e `Sensitive=True`. Essa tag do LF é usada para descrever o `col_tag_database` e a tabela `source_data_col_lvl` (em nível de coluna) quando conectada como `lf-data-analyst` no Amazon Athena.

a. Faça login no console do Lake Formation em https://console.aws.amazon.com/lakeformation/`lf-data-engineer`.

b. Na página Bancos de dados, selecione o banco de dados `col_tag_database`.

c. Selecione Ações e Concessão.

d. Em Entidades principais, selecione Usuários e perfis do IAM.

e. Para Usuários e perfis do IAM, selecione `lf-data-analyst`.

f. Selecione Recursos correspondentes às tags do LF.

g. Selecione Adicionar tag do LF.

h. Para Chave, selecione `Confidential`.

i. Para Valores, selecione `False`.

j. Selecione Adicionar tag do LF.

k. Para Chave, selecione `Sensitive`.

l. Para Valores, selecione `True`.

m. Para Permissões de banco de dados, selecione `Describe`.

n. Para Permissões de tabela, selecione `Select` e `Describe`.

o. Selecione Conceder.

Etapa 5: Executar uma consulta no Amazon Athena para verificar as permissões

Para essa etapa, use o Amazon Athena para executar consultas `SELECT` nas duas tabelas (`source_data and source_data_col_lvl`). Use o caminho do Amazon S3 como o local do resultado da consulta (`s3://lf-tagbased-demo-`*`Account-ID`*`/athena-results/`).

1. Faça login no console Athena em https://console.aws.amazon.com/athena/. `lf-data-analyst`
2. No editor de consultas do Athena, selecione `tag_database` no painel esquerdo.
3. Selecione o ícone de opções de menu adicionais (três pontos verticais) ao lado de `source_data` e selecione Exibir tabela.
4. Selecione Executar consulta.

A consulta deve levar alguns minutos para ser executada. A consulta exibe todas as colunas na saída porque a tag do LF está associada no nível do banco de dados e a tabela `source_data` a herdou automaticamente `LF-tag` do banco de dados `tag_database`.

5. Execute outra consulta usando `col_tag_database` e `source_data_col_lvl`.

   A segunda consulta retorna as duas colunas que foram marcadas como `Non-Confidential` e `Sensitive`.

6. Você também pode verificar o comportamento da política de acesso baseada em tags do Lake Formation em colunas para as quais você não tem concessões de políticas. Quando uma coluna não marcada é selecionada na tabela `source_data_col_lvl`, o Athena retorna um erro. Por exemplo, você pode executar a seguinte consulta para escolher colunas não marcadas `geolocationid`:

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

Etapa 6: limpar AWS os recursos

Para evitar cobranças indesejadas Conta da AWS, você pode excluir os AWS recursos usados neste tutorial.

1. Faça login no console do Lake Formation como `lf-data-engineer` e exclua os bancos de dados `tag_database` e `col_tag_database`.
2. Em seguida, faça login como `lf-data-steward` e limpe todas as permissões de tags do LF, Permissões de dados e Permissões de localização de dados que foram concedidas acima e que foram concedidas a `lf-data-engineer` e `lf-data-analyst`..
3. Faça login no console do Amazon S3 como proprietário da conta usando as credenciais do IAM que você usou para implantar a pilha. CloudFormation
4. Exclua os seguintes buckets:

   - lf-tagbased-demo-accesslogs-*acct-id*
   - lf-tagbased-demo-*acct-id*

5. Faça login no CloudFormation console em https://console.aws.amazon.com/cloudformation e exclua a pilha que você criou. Aguarde até que o status da pilha mude para `DELETE_COMPLETE`.

Como proteger os data lakes com controle de acesso em nível de linha

AWS Lake Formation as permissões em nível de linha permitem que você forneça acesso a linhas específicas em uma tabela com base nas políticas de governança e conformidade de dados. Se você tem tabelas grandes armazenando bilhões de registros, você precisa de uma forma de permitir que diferentes usuários e equipes acessem somente os dados que eles têm permissão para ver. O controle de acesso em nível de linha é uma maneira simples e eficaz de proteger os dados e, ao mesmo tempo, dar aos usuários acesso aos dados de que precisam para realizar seu trabalho. O Lake Formation fornece relatórios centralizados de auditoria e conformidade, identificando quais entidades principais acessaram quais dados, quando, e por meio de quais serviços.

Neste tutorial, você aprenderá como os controles de acesso em nível de linha funcionam no Lake Formation e como configurá-los.

Este tutorial inclui um AWS CloudFormation modelo para configurar rapidamente os recursos necessários. É possível revisá-lo e personalizá-lo para atender às suas necessidades.

Tópicos

- Público-alvo
- Pré-requisitos
- Etapa 1: Provisionar os recursos
- Etapa 2: Consulta sem filtros de dados
- Etapa 3: Configurar filtros de dados e conceder permissões
- Etapa 4: Consulta com filtros de dados
- Etapa 5: limpar AWS os recursos

Público-alvo

Este tutorial é destinado a administradores de dados, engenheiros de dados e analistas de dados. A tabela a seguir lista perfis e responsabilidades de um proprietário e um consumidor de dados.

Perfil	Description
Administrador do IAM	Um usuário que pode criar usuários e perfis e buckets do Amazon Simple Storage Service

Perfil	Description
	(Amazon S3). Tem a política `AdministratorAccess` AWS gerenciada.
Administrador do data lake	Um usuário responsável por configurar o data lake, criar filtros de dados e conceder permissões aos analistas de dados.
Analista de dados	Um usuário que pode executar consultas no data lake. Analistas de dados residentes em países diferentes (para nosso caso de uso, EUA e Japão) só podem analisar avaliações de produtos de clientes localizados em seus próprios países e, por motivos de conformidade, não devem ser capazes de ver dados de clientes localizados em outros países.

Pré-requisitos

Antes de começar este tutorial, você deve ter um Conta da AWS que possa ser usado para entrar como usuário administrativo com as permissões corretas. Para obter mais informações, consulte Inicial completa AWS tarefas de configuração.

O tutorial pressupõe que você esteja familiarizado com o IAM. Para obter informações sobre o IAM, consulte o Guia do usuário do IAM.

Alterar as configurações do Lake Formation

> **Important**
>
> Antes de iniciar o CloudFormation modelo, desative a opção Usar somente o controle de acesso do IAM para novos bancos de dados/tabelas no Lake Formation seguindo as etapas abaixo:

1. Faça login no console do Lake Formation https://console.aws.amazon.com/lakeformation/na região Leste dos EUA (Norte da Virgínia) ou Oeste dos EUA (Oregon).

2. Em catálogo de dados, selecione Configurações.
3. Desmarque Usar somente controle de acesso IAM para novos bancos de dados e Usar somente controle de acesso IAM para novas tabelas em novos bancos de dados.
4. Escolha Salvar.

Etapa 1: Provisionar os recursos

Este tutorial inclui um CloudFormation modelo para uma configuração rápida. É possível revisá-lo e personalizá-lo para atender às suas necessidades. O CloudFormation modelo gera os seguintes recursos:

- Usuários e políticas para:
  - DataLakeAdmin
  - DataAnalystEUA
  - DataAnalystJP
- Configurações e permissões do data lake do Lake Formation
- Uma função Lambda (para recursos CloudFormation personalizados apoiados pelo Lambda) usada para copiar arquivos de dados de amostra do bucket público do Amazon S3 para o seu bucket do Amazon S3
- Um bucket do Amazon S3 para servir como Data Lake.
- Um AWS Glue Data Catalog banco de dados, tabela e partição

Criar seus recursos

Siga estas etapas para criar seus recursos usando o CloudFormation modelo.

1. Faça login no CloudFormation console em https://console.aws.amazon.com/cloudformation na região Leste dos EUA (Norte da Virgínia).
2. Selecione Iniciar Pilha.
3. Na página Criar pilha, selecione Avançar.
4. Digite um Nome de pilha.
5. Para DatalakeAdminUserNamee DatalakeAdminUserPassword, insira seu nome de usuário e senha do IAM para o usuário administrador do data lake.

6. Para DataAnalystUsUserNamee DataAnalystUsUserPassword, insira o nome de usuário e a senha do nome de usuário e senha que você deseja para o usuário analista de dados responsável pelo mercado dos EUA.
7. DataAnalystJpUserPasswordEm DataAnalystJpUserNamee, insira o nome de usuário e a senha do nome de usuário e senha que você deseja para o usuário analista de dados responsável pelo mercado japonês.
8. Para DataLakeBucketName, insira o nome do seu repositório de dados.
9. Para DatabaseNamee TableNamedeixe como padrão.
10. Selecione Avançar.
11. Na página seguinte, selecione Avançar.
12. Analise os detalhes na página final e selecione Eu reconheço que isso CloudFormation pode criar recursos do IAM.
13. Escolha Criar.

    A criação da pilha pode levar um minuto para ser concluída.

Etapa 2: Consulta sem filtros de dados

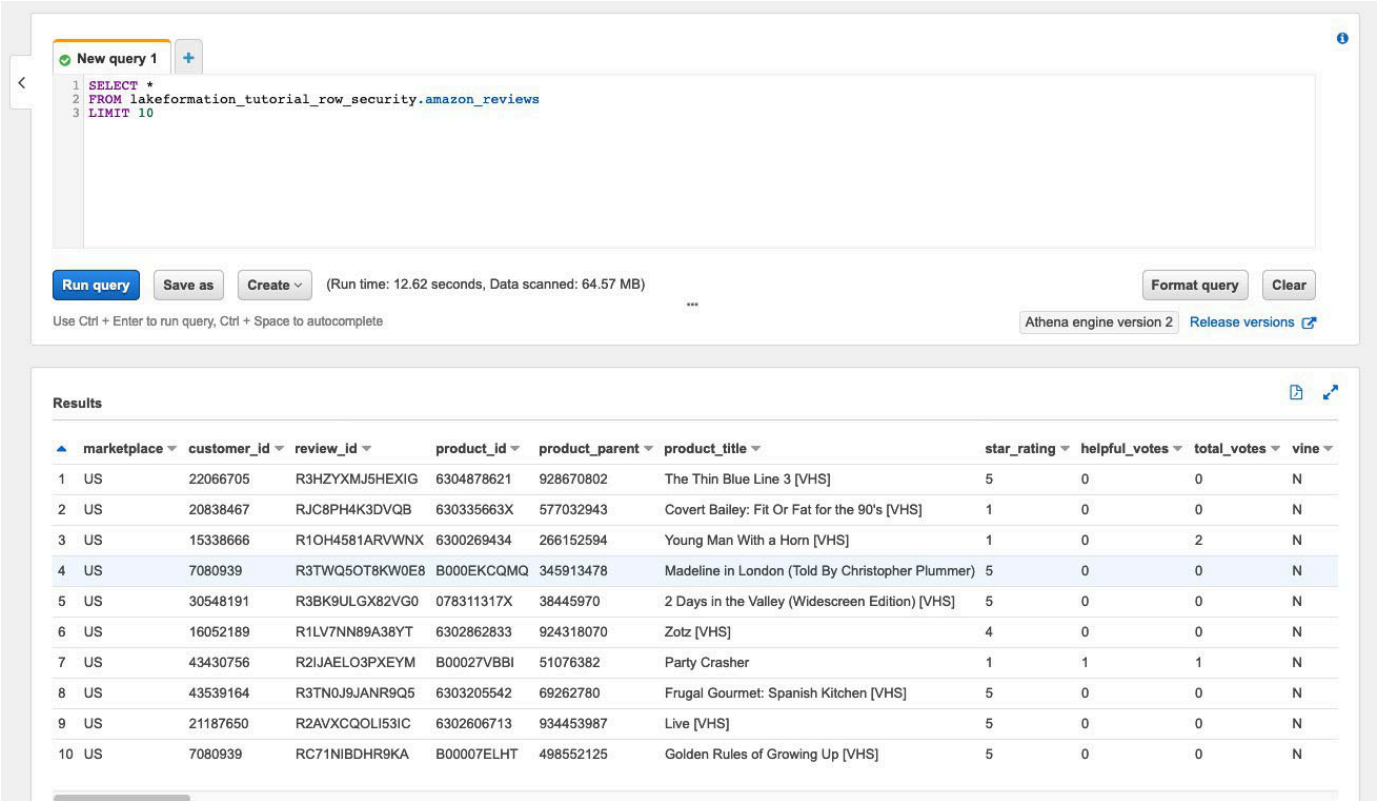
Depois que você configurar o ambiente, poderá consultar a tabela de avaliações de produtos. Primeiro, consulte a tabela sem controles de acesso em nível de linha para garantir que você possa ver os dados. Se você estiver executando consultas no Amazon Athena pela primeira vez, precisará configurar a localização do resultado da consulta.

Consulte a tabela sem controle de acesso em nível de linha

1. Faça login no Athena console https://console.aws.amazon.com/athena/como `DatalakeAdmin` usuário e execute a seguinte consulta:

   ```
   SELECT *
   FROM lakeformation_tutorial_row_security.amazon_reviews
   LIMIT 10
   ```

   A captura de tela a seguir mostra o resultado da consulta. Essa tabela tem apenas uma partição, `product_category=Video`, portanto, cada registro é um comentário de avaliação de um produto de vídeo.

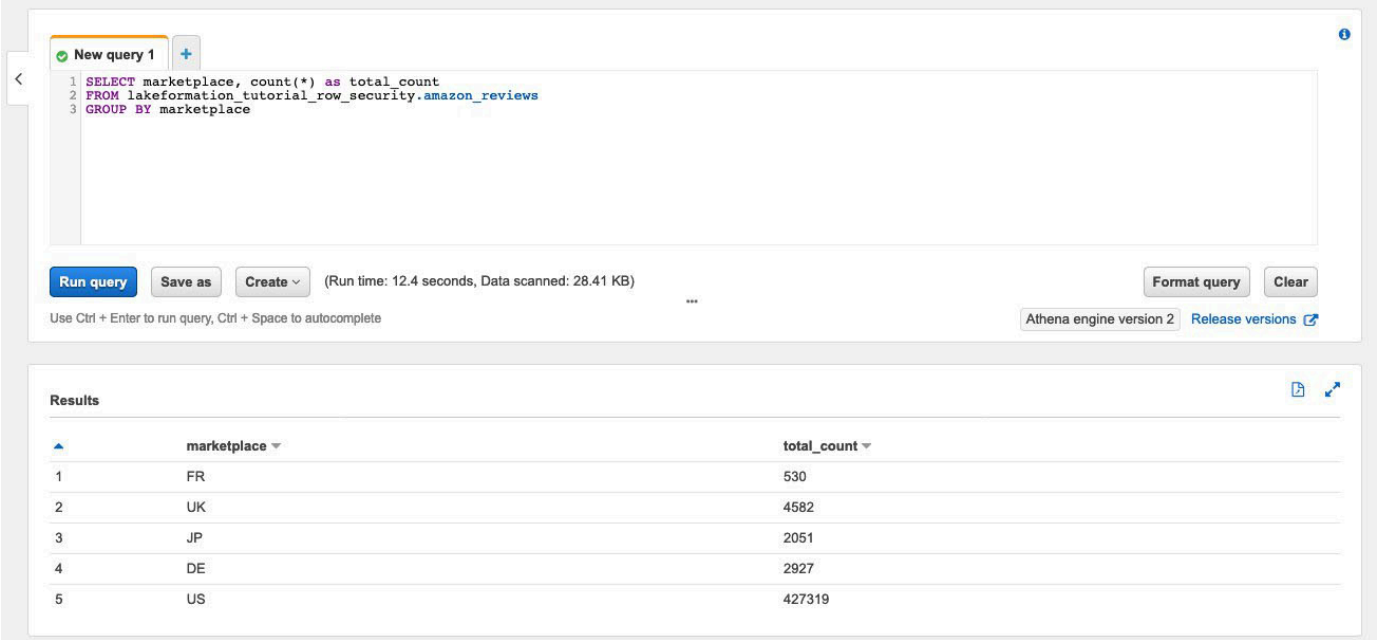


	marketplace	customer_id	review_id	product_id	product_parent	product_title	star_rating	helpful_votes	total_votes	vine
1	US	22066705	R3HZYXMJ5HEXIG	6304878621	928670802	The Thin Blue Line 3 [VHS]	5	0	0	N
2	US	20838467	RJC8PH4K3DVQB	630335663X	577032943	Covert Bailey: Fit Or Fat for the 90's [VHS]	1	0	0	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. Em seguida, execute uma consulta de agregação para recuperar o número total de registros por `marketplace`.

```
SELECT marketplace, count(*) as total_count
FROM lakeformation_tutorial_row_security.amazon_reviews
GROUP BY marketplace
```

A captura de tela a seguir mostra o resultado da consulta. A coluna `marketplace` tem cinco valores diferentes. Nas etapas subsequentes, você configurará filtros baseados em linhas usando a coluna `marketplace`.



Etapa 3: Configurar filtros de dados e conceder permissões

Este tutorial usa dois analistas de dados: um responsável pelo mercado dos EUA e outro pelo mercado japonês. Cada analista usa o Athena para analisar avaliações de clientes somente para seu mercado específico. Crie dois filtros de dados diferentes, um para o analista responsável pelo mercado dos EUA e outro para o responsável pelo mercado japonês. Em seguida, conceda aos analistas suas respectivas permissões.

Crie filtros de dados e conceda permissões

1. Crie um filtro para restringir o acesso aos dados do `marketplace` dos US.

   a. Faça login no console do Lake Formation https://console.aws.amazon.com/lakeformation/na região Leste dos EUA (Norte da Virgínia) como `DatalakeAdmin` usuário.

   b. Selecione Filtros de dados.

   c. Selecione Criar novo filtro.

   d. Em Nome do filtro de dados, digite `amazon_reviews_US`.

   e. Em Banco de dados de destino, selecione o banco de dados `lakeformation_tutorial_row_security`.

   f. Em Tabela de destino, selecione a tabela `amazon_reviews`.

   g. Para acesso em nível de coluna, deixe como padrão.

h. Em Expressão de filtro de linha, digite `marketplace='US'`.

i. Selecione Criar filtro.

2. Crie um filtro para restringir o acesso aos dados japoneses do `marketplace`.

a. Na página Filtros de dados, selecione Criar novo filtro.

b. Em Nome do filtro de dados, digite `amazon_reviews_JP`.

c. Em Banco de dados de destino, selecione o banco de dados `lakeformation_tutorial_row_security`.

d. Em Tabela de destino, selecione o `table amazon_reviews`.

e. Para acesso em nível de coluna, deixe como padrão.

f. Em Expressão de filtro de linha, digite `marketplace='JP'`.

g. Selecione Criar filtro.

3. Em seguida, conceda permissões aos analistas de dados usando esses filtros de dados. Siga estas etapas para conceder permissões ao analista de dados dos EUA (`DataAnalystUS`):

a. Em Permissões, escolha Permissões do Data lake.

b. Em Permissão de dados, selecione Conceder.

c. Para Entidades principais, selecione Usuários e perfis do IAM e selecione o perfil `DataAnalystUS`.

d. Em Tags do LF ou recursos de catálogo, selecione Recursos nomeados de catálogo de dados.

e. Para o Banco de dados, selecione `lakeformation_tutorial_row_security`.

f. Para Tabelas opcionais, selecione `amazon_reviews`.

g. Para Filtros de dados — opcional, selecione `amazon_reviews_US`.

h. Para Permissões de filtro de dados, marque Selecionar.

i. Selecione Conceder.

4. Siga estas etapas para conceder permissões ao analista de dados japonês (`DataAnalystJP`):

a. Em Permissões, escolha Permissões do Data lake.

b. Em Permissão de dados, selecione Conceder.

c. Para Entidades principais, selecione Usuários e perfis do IAM e selecione o perfil `DataAnalystJP`.

d. Em Tags do LF ou recursos de catálogo, selecione Recursos nomeados de catálogo de dados.

e. Para o Banco de dados, selecione `lakeformation_tutorial_row_security`.

f. Para Tabelas opcionais, selecione `amazon_reviews`.

g. Para Filtros de dados — opcional, selecione `amazon_reviews_JP`.

h. Para Permissões de filtro de dados, marque Selecionar.

i. Selecione Conceder.

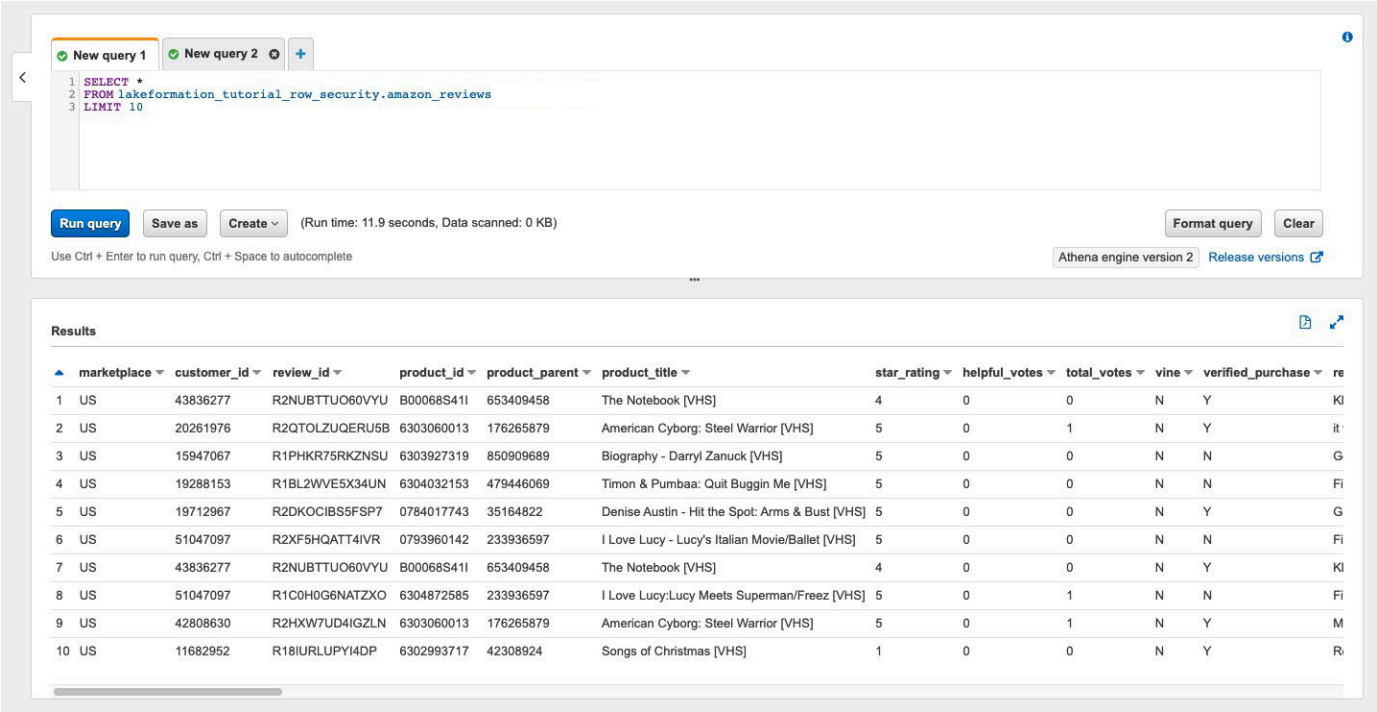
Etapa 4: Consulta com filtros de dados

Com os filtros de dados anexados à tabela de avaliações de produtos, faça algumas consultas e veja como as permissões são aplicadas pelo Lake Formation.

1. Faça login no console do Athena https://console.aws.amazon.com/athena/como usuário. `DataAnalystUS`
2. Execute a consulta a seguir para recuperar alguns registros, que são filtrados com base nas permissões em nível de linha que definimos:

```
SELECT *
FROM lakeformation_tutorial_row_security.amazon_reviews
LIMIT 10
```

A captura de tela a seguir mostra o resultado da consulta.



	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. Da mesma forma, execute uma consulta para contar o número total de registros por loja.

```
SELECT marketplace , count ( * ) as total_count
FROM lakeformation_tutorial_row_security .amazon_reviews
GROUP BY marketplace
```

O resultado da consulta mostra apenas o US dos `marketplace` nos resultados. Isso ocorre porque o usuário só pode ver as linhas em que o valor da coluna `marketplace` é igual a US.

4. Mude para o usuário `DataAnalystJP` e execute a mesma consulta.

```
SELECT *
FROM lakeformation_tutorial_row_security.amazon_reviews
LIMIT 10
```

O resultado da consulta mostra apenas os registros que pertencem ao domínio `JP marketplace`.

5. Execute a consulta para contar o número total de registros por `marketplace`.

```
SELECT marketplace, count(*) as total_count
FROM lakeformation_tutorial_row_security.amazon_reviews
GROUP BY marketplace
```

O resultado da consulta mostra apenas a linha pertencente ao domínio `JP` `marketplace`.

Etapa 5: limpar AWS os recursos

Limpar recursos

Para evitar cobranças indesejadas Conta da AWS, você pode excluir os AWS recursos usados neste tutorial.

- Exclua a pilha do Cloud Formation.

Compartilhamento de um Data Lake usando controle de acesso baseado em tags do Lake Formation e recursos nomeados

Este tutorial demonstra como você pode configurar AWS Lake Formation para compartilhar com segurança os dados armazenados em um data lake com várias empresas, organizações ou unidades de negócios, sem precisar copiar o banco de dados inteiro. Há duas opções para compartilhar seus bancos de dados e tabelas com outra pessoa Conta da AWS usando o controle de acesso entre contas do Lake Formation:

- Controle de acesso baseado em tags do Lake Formation (LF-TBAC) (recomendado)

  O controle de acesso baseado em tags do Lake Formation é uma estratégia de autorização que define permissões com base em atributos. No Lake Formation, esses atributos são chamados de tags do LF. Para obter mais detalhes, consulte Como gerenciar um data lake usando o controle de acesso baseado em tags do Lake Formation.
- Recursos nomeados do Lake Formation

  O método de recurso chamado Lake Formation é uma estratégia de autorização que define permissões para recursos. Os recursos incluem bancos de dados, tabelas e colunas. Os administradores do Data Lake podem atribuir e revogar permissões em recursos do Lake Formation. Para obter mais detalhes, consulte Cross-account compartilhamento de dados em Lake Formation.

  Recomendamos usar recursos nomeados se o administrador do Data Lake preferir conceder permissões explicitamente a recursos individuais. Quando você usa o método de recurso nomeado para conceder permissões do Lake Formation em um recurso do Catálogo de Dados para

uma conta externa, o Lake Formation usa AWS Resource Access Manager (AWS RAM) para compartilhar o recurso.

Tópicos

- Público-alvo
- Definir as configurações do catálogo de dados do Lake Formation na conta do produtor
- Etapa 1: provisionar seus recursos usando AWS CloudFormation modelos
- Etapa 2: Pré-requisitos de compartilhamento entre contas do Lake Formation
- Etapa 3: Implementar o compartilhamento entre contas usando o método de controle de acesso baseado em tags
- Etapa 4: Implementar o método de recurso nomeado
- Etapa 5: limpar AWS os recursos

Público-alvo

Este tutorial é destinado a administradores de dados, engenheiros de dados e analistas de dados. Quando se trata de compartilhar tabelas do Catálogo de Dados AWS Glue e administrar permissões no Lake Formation, os administradores de dados nas contas produtoras têm propriedade funcional com base nas funções que suportam e podem conceder acesso a vários consumidores, organizações externas e contas. A tabela a seguir lista os perfis usados neste tutorial:

Perfil	Description
DataLakeAdminProducer	O usuário do IAM administrador do Data Lake tem o seguinte acesso: • Acesso completo de leitura, gravação e atualização a todos os recursos do catálogo de dados • Capacidade de conceder permissões aos recursos • Criar um link de recurso para a tabela compartilhada

Perfil	Description
	• Pode anexar tags do LF aos recursos, fornecendo acesso às entidades principais com base em quaisquer políticas criadas por administradores de dados
DataLakeAdminConsumer	O usuário do IAM administrador do Data Lake tem o seguinte acesso: • Acesso completo de leitura, gravação e atualização a todos os recursos do catálogo de dados • Capacidade de conceder permissões aos recursos • Criar um link de recurso para a tabela compartilhada • Pode anexar tags do LF aos recursos, fornecendo acesso às entidades principais com base em quaisquer políticas criadas por administradores de dados
DataAnalyst	O DataAnalyst usuário tem o seguinte acesso: • Acesso refinado a recursos compartilhados pelas políticas de acesso baseadas em tags do Lake Formation ou usando o método de recursos nomeados

Definir as configurações do catálogo de dados do Lake Formation na conta do produtor

Antes de começar este tutorial, você deve ter um Conta da AWS que possa ser usado para entrar como usuário administrativo com as permissões corretas. Para obter mais informações, consulte Inicial completa AWS tarefas de configuração.

O tutorial pressupõe que você esteja familiarizado com o IAM. Para obter informações sobre o IAM, consulte o Guia do usuário do IAM.

Definir as configurações do catálogo de dados do Lake Formation na conta do produtor

> **Note**
>
> Neste tutorial, a conta que tem a tabela de origem é chamada de conta do produtor, e a conta que precisa acessar a tabela de origem é chamada de conta do consumidor.

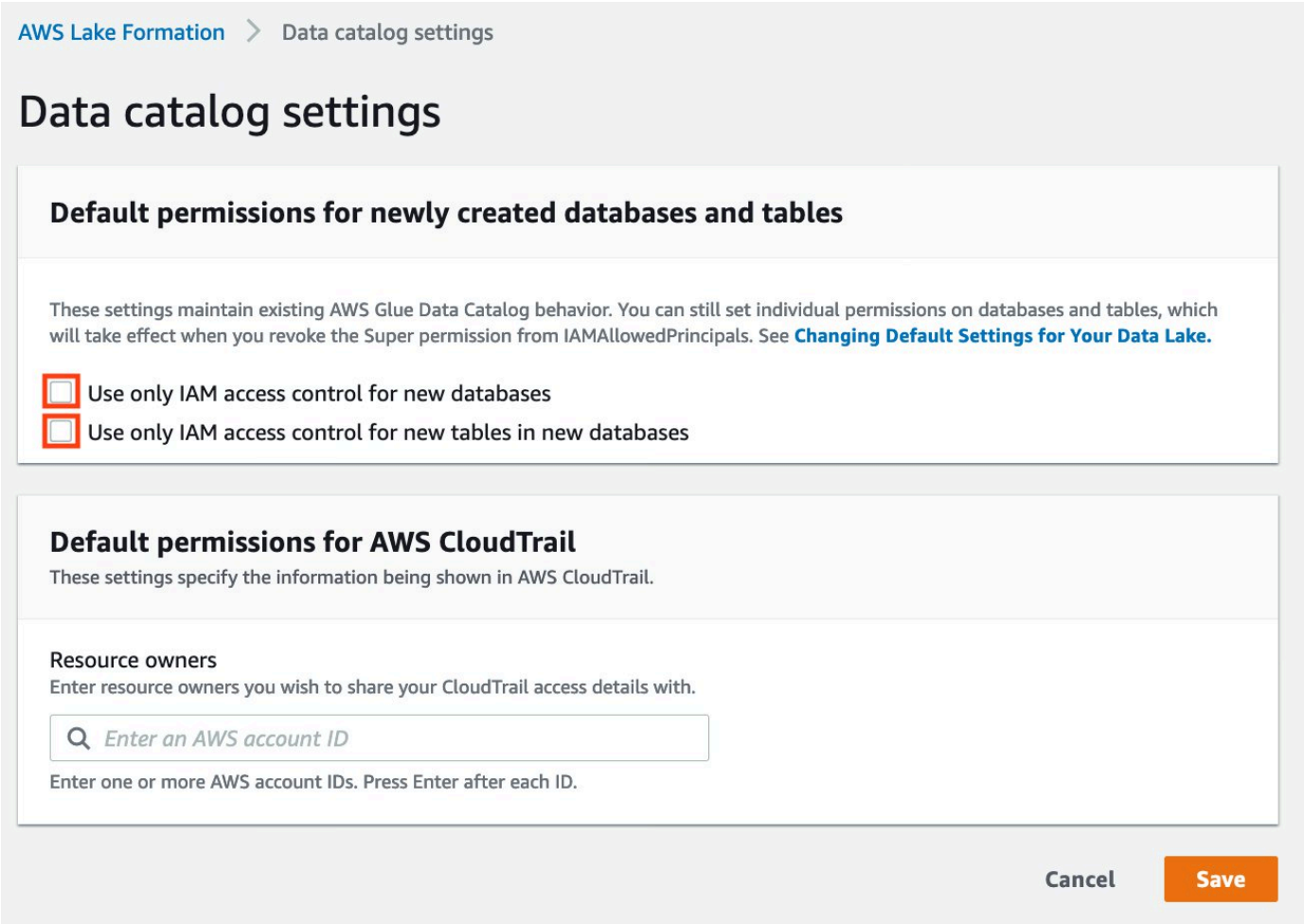
O Lake Formation fornece seu próprio modelo de gerenciamento de permissões. Para manter a compatibilidade com versões anteriores do modelo de permissão do IAM, a `Super` permissão é concedida ao grupo `IAMAllowedPrincipals` em todos os AWS Glue Data Catalog recursos existentes por padrão. Além disso, a opção Usar somente as configurações de controle de acesso do IAM estão habilitadas para novos recursos do catálogo de dados. Este tutorial usa controle de acesso refinado usando permissões do Lake Formation e usa políticas do IAM para controle de acesso de único fator. Para mais detalhes, consulte Métodos para controle de acesso de alta granularidade. Portanto, antes de usar um AWS CloudFormation modelo para uma configuração rápida, você precisa alterar as configurações do Lake Formation Data Catalog na conta do produtor.

> **Important**
>
> Essa configuração afeta todos os bancos de dados e tabelas recém-criados, portanto, é altamente recomendável concluir este tutorial em uma conta que não seja de produção ou em uma nova conta. Além disso, se você estiver usando uma conta compartilhada (como a conta de desenvolvimento da sua empresa), certifique-se de que ela não afete outros recursos. Se você preferir manter as configurações de segurança padrão, conclua uma etapa adicional ao compartilhar recursos com outras contas, na qual você revoga a Super permissão padrão do banco de dados ou da tabela do `IAMAllowedPrincipals`. Entraremos em detalhes mais adiante neste tutorial.

Para definir as configurações do catálogo de dados do Lake Formation na conta do produtor, conclua as etapas a seguir:

1. Faça login Console de gerenciamento da AWS usando a conta do produtor como usuário administrador ou como usuário com permissão da `PutDataLakeSettings` API Lake Formation.
2. No console do Lake Formation, no painel de navegação, em catálogo de dados, selecione Configurações.
3. Desmarque Usar somente controle de acesso IAM para novos bancos de dados e Usar somente controle de acesso IAM para novas tabelas em novos bancos de dados

   Escolha Salvar.



   Além disso, você pode remover as permissões de `CREATE_DATABASE` para `IAMAllowedPrincipals` em Perfis e tarefas administrativas, Criadores de banco de dados. Só então, você pode determinar quem pode criar um novo banco de dados por meio das permissões do Lake Formation.

Etapa 1: provisionar seus recursos usando AWS CloudFormation modelos

O CloudFormation modelo da conta do produtor gera os seguintes recursos:

- Um bucket do Amazon S3 para servir como Data Lake.
- Uma função Lambda (para recursos personalizados apoiados pelo Lambda CloudFormation). Usamos a função para copiar exemplos de arquivos de dados do bucket público do Amazon S3 para o bucket do Amazon S3.
- Usuários e políticas do IAM: DataLakeAdminProducer.
- As configurações e permissões apropriadas do Lake Formation, incluindo:
  - Definir o administrador do data lake do Lake Formation na conta do produtor
  - Registrar um bucket do Amazon S3 como local do Data Lake do Lake Formation (conta do produtor)
- Um AWS Glue Data Catalog banco de dados, tabela e partição. Como há duas opções para compartilhar recursos Contas da AWS, esse modelo cria dois conjuntos separados de banco de dados e tabela.

O CloudFormation modelo da conta do consumidor gera os seguintes recursos:

- Usuários e políticas do IAM:
  - DataLakeAdminConsumer
  - DataAnalyst
- Um AWS Glue Data Catalog banco de dados. Esse banco de dados serve para criar links de recursos para recursos compartilhados.

Crie seus recursos na conta do produtor

1. Faça login no AWS CloudFormation console em https://console.aws.amazon.com/cloudformation na região Leste dos EUA (Norte da Virgínia).
2. Selecione Iniciar Pilha.
3. Selecione Avançar.
4. Em Nome da pilha, digite um nome de pilha, como `stack-producer`.
5. Na seção Configuração do usuário, insira o nome de usuário e a senha para `ProducerDatalakeAdminUserName` e `ProducerDatalakeAdminUserPassword`.

6. Para DataLakeBucketName, insira o nome do seu bucket do data lake. Esse nome precisa ser globalmente exclusivo.
7. Para DatabaseNamee TableName, deixe os valores padrão.
8. Escolha Próximo.
9. Na página seguinte, selecione Avançar.
10. Analise os detalhes na página final e selecione Eu reconheço que isso AWS CloudFormation pode criar recursos do IAM.
11. Escolha Criar.

    A criação da pilha pode levar até um minuto.

Crie seus recursos na conta do consumidor

1. Faça login no AWS CloudFormation console em https://console.aws.amazon.com/cloudformation na região Leste dos EUA (Norte da Virgínia).
2. Selecione Iniciar Pilha.
3. Selecione Avançar.
4. Em Nome da pilha, digite um nome de pilha, como `stack-consumer`.
5. Na seção Configuração do usuário, insira o nome de usuário e a senha para `ConsumerDatalakeAdminUserName` e `ConsumerDatalakeAdminUserPassword`.
6. Para `DataAnalystUserName` e `DataAnalystUserPassword`, insira o nome de usuário e a senha que você deseja para o usuário do IAM do analista de dados.
7. Para DataLakeBucketName, insira o nome do seu bucket do data lake. Esse nome precisa ser globalmente exclusivo.
8. Para DatabaseName, deixe os valores padrão.
9. Em `AthenaQueryResultS3BucketName`, insira o nome do bucket do Amazon S3 que armazena resultados da consulta do Amazon Athena. Se você não tiver um, crie um bucket da Amazon S3.
10. Selecione Avançar.
11. Na página seguinte, selecione Avançar.
12. Analise os detalhes na página final e selecione Eu reconheço que isso AWS CloudFormation pode criar recursos do IAM.
13. Escolha Criar.

A criação da pilha pode levar até um minuto.

> **Note**
>
> Depois de concluir o tutorial, exclua a pilha CloudFormation para evitar cobranças. Verifique se os recursos foram excluídos com sucesso no status de eventos da pilha.

Etapa 2: Pré-requisitos de compartilhamento entre contas do Lake Formation

Antes de compartilhar recursos com o Lake Formation, há pré-requisitos tanto para o método de controle de acesso baseado em tags quanto para o método de recursos nomeados.

Controle de acesso completo com base em tags: pré-requisitos do para compartilhamento de dados do entre contas

- Para obter mais informações sobre os requisitos de compartilhamento de dados entre contas, consulte a seção Pré-requisitos no capítulo Compartilhamento de dados entre contas.

  Para compartilhar recursos do Catálogo de Dados com a versão 3 ou superior das configurações da versão Cross Account, o concedente precisa ter as permissões do IAM definidas na política AWS gerenciada `AWSLakeFormationCrossAccountManager` em sua conta.

  Se você estiver usando a versão 1 ou a versão 2 das Configurações de versão entre contas, antes de usar o método de controle de acesso baseado em tags para conceder acesso entre contas aos recursos, você deve adicionar o seguinte objeto de permissões `JSON` à política de recursos do catálogo de dados na conta do produtor. Isso dá à conta do consumidor permissão para acessar o catálogo de dados quando `glue:EvaluatedByLakeFormationTags` é verdadeiro. Além disso, essa condição se torna verdadeira para recursos nos quais você concedeu permissão usando tags de permissão do Lake Formation para a conta do consumidor. Essa política é necessária Conta da AWS para todas as quais você está concedendo permissões.

  A política a seguir deve estar dentro de um elemento`Statement`. Discutiremos a Política do IAM completa na próxima seção.

```
{
    "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
        }
    }
}
```

Preencha os pré-requisitos de compartilhamento entre contas do método de recurso nomeado

1. Se não houver uma política de recursos do catálogo de dados em sua conta, as concessões entre contas do Lake Formation que você fizer procederão normalmente. No entanto, se existir uma política de recursos do catálogo de dados, você deverá adicionar a seguinte declaração a ela para permitir que suas concessões entre contas sejam bem-sucedidas se forem feitas com o método de recurso nomeado. Se você planeja usar somente o método de recurso nomeado ou somente o método de controle de acesso baseado em tag, você pode pular esta etapa. Neste tutorial, avaliamos os dois métodos e precisamos adicionar a política a seguir.

   A política a seguir deve estar dentro de um elemento`Statement`. Discutiremos a Política do IAM completa na próxima seção.

   ```
   {
             "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. Em seguida, adicione a política de AWS Glue Data Catalog recursos usando o AWS Command Line Interface (AWS CLI).

   Se você conceder permissões entre contas usando o método de controle de acesso baseado em tags e o método de recurso nomeado, defina o argumento `EnableHybrid` como “true” ao adicionar as políticas anteriores. Como essa opção não é compatível com o console no momento, é necessário usar a API `glue:PutResourcePolicy` e a AWS CLI.

   Primeiro, crie um documento de política (como policy.json) e adicione as duas políticas anteriores. *`consumer-account-id`*Substitua pela *`account ID`* do Conta da AWS destinatário da concessão, *`region`* pela região do catálogo de dados contendo os bancos de dados e tabelas nos quais você está concedendo permissões e *`account-id`* pela Conta da AWS ID do produtor.

   Digite o AWS CLI comando a seguir. *`glue-resource-policy`*Substitua pelos valores corretos (como file: //policy.json).

   ```
   aws glue put-resource-policy --policy-in-json glue-resource-policy --enable-hybrid
    TRUE
   ```

   Para obter mais informações, consulte put-resource-policy.

Etapa 3: Implementar o compartilhamento entre contas usando o método de controle de acesso baseado em tags

Nesta seção, orientamos sobre estas etapas de alto nível:

1. Defina uma tag do LF.

2. Atribua a tag do LF ao recurso de destino.
3. Conceda permissão da tag do LF à conta do consumidor.
4. Conceda permissão de dados à conta do consumidor.
5. Opcionalmente, revogue as permissões para `IAMAllowedPrincipals` no banco de dados, nas tabelas e nas colunas.
6. Crie um link de recurso para a tabela compartilhada.
7. Crie uma tag do LF e atribua-a ao banco de dados de destino.
8. Conceda permissão de dados da tag do LF à conta do consumidor.

Defina uma tag do LF

> **Note**
>
> Se você estiver conectado à sua conta de produtor, saia antes de concluir as etapas a seguir.

1. Faça login na conta do produtor como administrador do data lake em https://console.aws.amazon.com/lakeformation/. Use o número da conta do produtor, o nome de usuário do IAM (o padrão é `DatalakeAdminProducer`) e a senha que você especificou durante a criação da pilha do CloudFormation .
2. No console do Lake Formation (https://console.aws.amazon.com/lakeformation/), no painel de navegação, em Permissões, escolha LF-Tags and Permissions.
3. Selecione Adicionar tag do LF.

Atribua a tag do LF ao recurso de destino

Atribua a tag do LF ao recurso de destino e conceda permissões de dados para outra conta

Como administrador do Data Lake, você pode anexar tags aos recursos. Se você planeja usar uma função separada, talvez seja necessário conceder permissões, descrever e anexar ao perfil separado.

1. No painel de navegação, em catálogo de dados, selecione Bancos de Dados.
2. Selecione o banco de dados de destino (`lakeformation_tutorial_cross_account_database_tbac`) e, no menu Ações, selecione Editar tags do LF.

Neste tutorial, você atribui uma tag do LF a um banco de dados, mas também pode atribuir tags do LF a tabelas e colunas.

3. Escolha Atribuir nova tag do LF.
4. Adicione a chave `Confidentiality` e o valor `public`.
5. Selecione Salvar.

Conceda permissão da tag do LF à conta do consumidor.

Ainda na conta do produtor, conceda permissões à conta do consumidor para acessar a tag do LF.

1. No painel de navegação, em Permissões, selecione permissões e tags do LF.
2. Na guia Tags do LF, selecione a chave e os valores da tag do LF que está sendo compartilhada com a conta do consumidor (chave `Confidentiality` e valor `public`).
3. Escolha Conceder permissões.
4. Em Tipo de permissão, escolha Permissões do par de chave-valor da tag do LF.
5. Para Entidades principais, selecione Contas externas.
6. Insira o ID da Conta da AWS de destino.

   Contas da AWS dentro da mesma organização aparecem automaticamente. Caso contrário, você precisará inserir manualmente o Conta da AWS ID.
7. Em Permissões, selecione Descrever.

   Essas são permissões concedidas à conta do consumidor. Permissões concedidas são permissões que a conta do consumidor pode conceder a outras entidades principais.
8. Selecione Conceder.

   Nesse momento, o administrador do data Lake do consumidor deve conseguir encontrar a tag da política que está sendo compartilhada por meio do console do Lake Formation da conta do consumidor, em Permissões, Tags do LF e permissões.

Conceda a permissão de dados à conta do consumidor.

Agora, forneceremos acesso aos dados da conta do consumidor especificando uma expressão de tag do LF, concedendo à conta do consumidor acesso a qualquer tabela ou banco de dados que corresponda à expressão.

1. No painel de navegação, em Permissões, Permissões do Data Lake, selecione Conceder.
2. Em Diretores, escolha Contas externas e insira a Conta da AWS ID de destino.
3. Em Tags do LF ou recursos de catálogo, escolha a chave e os valores da tag do LF que está sendo compartilhada com a conta do consumidor (chave `Confidentiality` e valor `public`).
4. Em Permissões, em Recursos combinados com tags do LF (recomendado), selecione Adicionar tag do LF.
5. Selecione a chave e o valor da tag que está sendo compartilhada com a conta do consumidor (chave `Confidentiality` e valor `public`).
6. Para permissões de banco de dados, selecione Descrever em Permissões de banco de dados para conceder permissões de acesso no nível do banco de dados.
7. O administrador do data lake do consumidor deve ser capaz de encontrar a tag de política que está sendo compartilhada por meio da conta do consumidor no console do Lake Formation em https://console.aws.amazon.com/lakeformation/, em Permissões, funções e tarefas administrativas, LF-Tags.
8. Selecione Descrever em Permissões concedíveis para que a conta do consumidor possa conceder permissões em nível de banco de dados a seus usuários.
9. Para Permissões de tabela e coluna, selecione Selecionar e Descrever em Permissões de tabela.
10. Selecione Selecionar e Descrever em Permissões concedíveis.
11. Selecione Conceder.

Revogue a permissão para **`IAMAllowedPrincipals`** no banco de dados, nas tabelas e nas colunas (opcional).

No início deste tutorial, você alterou as configurações do Lake Formation Data Catalog. Se você pulou essa parte, esta etapa é necessária. Se tiver alterado as configurações do catálogo de dados do Lake Formation, ignore esta etapa.

Nesta etapa, precisamos revogar a Super permissão padrão do banco de dados `IAMAllowedPrincipals` ou da tabela. Para mais detalhes, consulte Etapa 4: mude seus armazenamentos de dados para o modelo de permissões do Lake Formation.

Antes de revogar a permissão de `IAMAllowedPrincipals`, certifique-se de ter concedido às entidades principais existentes do IAM a permissão necessária por meio do Lake Formation. Isso inclui três etapas:

1. Adicione a permissão do IAM ao usuário ou perfil do IAM de destino com a ação `GetDataAccess` do Lake Formation (com a política do IAM).
2. Conceda ao usuário ou perfil de destino do IAM com permissões de dados do Lake Formation (alterar, selecionar e assim por diante).
3. Em seguida, revogue as permissões para `IAMAllowedPrincipals`. Caso contrário, depois de revogar as permissões para `IAMAllowedPrincipals`, as entidades principais existentes do IAM talvez não consigam mais acessar o banco de dados ou o catálogo de dados de destino.

   A revogação da Super permissão para `IAMAllowedPrincipals` é necessária quando você deseja aplicar o modelo de permissão do Lake Formation (em vez do modelo de política do IAM) para gerenciar o acesso do usuário em uma única conta ou entre várias contas usando o modelo de permissão do Lake Formation. Você não precisa revogar a permissão de `IAMAllowedPrincipals` outras tabelas nas quais deseja manter o modelo de política tradicional do IAM.

   Nesse ponto, o administrador do data lake da conta do consumidor deve ser capaz de encontrar o banco de dados e a tabela que estão sendo compartilhados por meio da conta do consumidor no console do Lake Formation em https://console.aws.amazon.com/lakeformation/bancos de dados, em Catálogo de dados. Caso contrário, confirme se o seguinte está configurado corretamente:

   1. A tag e os valores de política corretos são atribuídos aos bancos de dados e tabelas de destino.
   2. A permissão correta de tag e a permissão de dados são atribuídas à conta do consumidor.
   3. Revogue a super permissão padrão do banco de dados `IAMAllowedPrincipals` ou da tabela.

Criar um link de recurso para a tabela compartilhada

Quando um recurso é compartilhado entre contas e os recursos compartilhados não são colocados no catálogo de dados das contas do consumidor. Para disponibilizá-los e consultar os dados subjacentes de uma tabela compartilhada usando serviços como o Athena, precisamos criar um link de recurso para a tabela compartilhada. Um link de recurso é um objeto do catálogo de dados que é um link para um banco de dados, ou uma tabela local ou compartilhada. Para obter detalhes, consulte Criação de links de recursos. Ao criar um link de recurso, você pode:

- Atribuir um nome diferente a um banco de dados ou tabela que esteja alinhado às políticas de nomenclatura de recursos do catálogo de dados.
- Usar serviços como Athena e Redshift Spectrum para consultar bancos de dados ou tabelas compartilhados.

Para criar um link de recurso, conclua as etapas a seguir:

1. Se você estiver conectado à sua conta de consumidor, saia.
2. Faça login como administrador do Data Lake da conta do consumidor. Use o ID da conta do consumidor, o nome de usuário do IAM (padrão DatalakeAdminConsumer) e a senha que você especificou durante a criação da CloudFormation pilha.
3. No console do Lake Formation (https://console.aws.amazon.com/lakeformation/), no painel de navegação, em Catálogo de dados, bancos de dados, escolha o banco de dados `lakeformation_tutorial_cross_account_database_tbac` compartilhado.

   Se você não vir o banco de dados, revise as etapas anteriores para ver se tudo está configurado corretamente.
4. Selecione Visualizar Tabelas.
5. Selecione a tabela compartilhada `amazon_reviews_table_tbac`.
6. No menu Ações, selecione Criar link de recurso.
7. Em Nome do link do recurso, digite um nome (para este tutorial, `amazon_reviews_table_tbac_resource_link`).
8. Em Banco de dados, selecione o banco de dados no qual o link do recurso foi criado (para esta postagem, a pilha CloudFormation n criou o banco de dados`lakeformation_tutorial_cross_account_database_consumer`).
9. Escolha Criar.

   O link do recurso aparece em catálogo de dados, Tabelas.

Crie uma tag do LF e atribua-a ao banco de dados de destino

As tags do Lake Formation residem no mesmo catálogo de dados dos recursos. Isso significa que as tags criadas na conta do produtor não estão disponíveis para uso ao conceder acesso aos links de recursos na conta do consumidor. Você precisa criar um conjunto separado de tags do LF na conta

do consumidor para usar o controle de acesso baseado em tags do LF ao compartilhar os links de recursos na conta do consumidor.

1. Defina a tag do LF na conta do consumidor. Para este tutorial, usamos a chave `Division` e os valores `sales`, `marketing` e `analyst`.
2. Atribua a chave da tag do LF `Division` e o valor `analyst` ao banco de dados `lakeformation_tutorial_cross_account_database_consumer`, onde o link de recurso é criado.

Conceda permissão de dados da tag do LF ao consumidor

Como etapa final, conceda permissão de dados da tag do LF ao consumidor.

1. No painel de navegação, em Permissões, Permissões do Data Lake, selecione Conceder.
2. Para Entidade principais, selecione Usuários e perfis do IAM e selecione o usuário `DataAnalyst`.
3. Em Recursos de tags do LF ou de catálogo, selecione Recursos correspondentes a tags do LF (recomendado).
4. Selecione Chave Divisão e analista de valor.
5. Para Permissões de banco de dados, selecione Descrever em Permissões de banco de dados.
6. Para Permissões de tabela e coluna, selecione Selecionar e Descrever em Permissões de tabela.
7. Selecione Conceder.
8. Repita essas etapas para o usuário `DataAnalyst`, onde a chave tag do LF é `Confidentiality` e o valor é `public`.

   Nesse ponto, o usuário analista de dados na conta do consumidor deve ser capaz de encontrar o link do banco de dados e do recurso e consultar a tabela compartilhada por meio do console do Athena em. https://console.aws.amazon.com/athena/ Caso contrário, confirme se o seguinte está configurado corretamente:

   - O link do recurso é criado para a tabela compartilhada
   - Você concedeu ao usuário acesso à tag do LF compartilhada pela conta do produtor
   - Você concedeu ao usuário acesso à tag do LF associada ao link do recurso e ao banco de dados no qual o link do recurso foi criado

- Verifique se você atribuiu a tag do LF correta ao link do recurso e ao banco de dados no qual o link do recurso foi criado

Etapa 4: Implementar o método de recurso nomeado

Para usar o método de recurso nomeado, orientamos você nas seguintes etapas de alto nível:

1. Opcionalmente, revogue a permissão para `IAMAllowedPrincipals` no banco de dados, nas tabelas e nas colunas.
2. Conceda a permissão de dados à conta do consumidor.
3. Aceite um compartilhamento de recursos de AWS Resource Access Manager.
4. Crie um link de recurso para a tabela compartilhada.
5. Conceda permissão de dados para a tabela compartilhada ao consumidor.
6. Conceda permissão de dados para o link de recurso ao consumidor.

Revogue a permissão para **`IAMAllowedPrincipals`** no banco de dados, nas tabelas e nas colunas (opcional).

- No início deste tutorial, alteramos as configurações do Lake Formation Data Catalog. Se você pulou essa parte, esta etapa é necessária. Para obter instruções, consulte a etapa opcional na seção anterior.

Conceda a permissão de dados à conta do consumidor.

1.
   > **Note**
   >
   > Se você estiver conectado à conta do produtor como outro usuário, saia primeiro.

   Faça login no console do Lake Formation https://console.aws.amazon.com/lakeformation/usando o administrador do data lake da conta do produtor usando o Conta da AWS ID, o nome de usuário do IAM (o padrão é`DatalakeAdminProducer`) e a senha especificados durante a criação da CloudFormation pilha.
2. Na página Permissões, em Permissões do Data Lake, selecione Conceder.

3. Em Diretores, escolha Contas externas e insira uma Conta da AWS IDs ou mais AWS organizações IDs. Para obter mais informações, consulte: Organizações AWS.

   As organizações às quais a conta do produtor pertence e Contas da AWS dentro da mesma organização aparecem automaticamente. Caso contrário, insira manualmente o ID da conta ou o ID da organização.
4. Em Tags do LF ou recursos de catálogo, selecione `Named data catalog resources`.
5. Em Bancos de dados, selecione o banco de dados `lakeformation_tutorial_cross_account_database_named_resource`.
6. Selecione Adicionar tag do LF.
7. Em Tabelas, selecione Todas as tabelas.
8. Para Permissões de coluna de tabela, selecione Selecionar e Descrever em Permissões de tabela.
9. Escolha Selecionar e Descrever em Permissões concedíveis.
10. Opcionalmente, para Permissões de dados, escolha Acesso simples baseado em coluna se for necessário o gerenciamento de permissões em nível de coluna.
11. Selecione Conceder.

Se você não revogou a permissão para `IAMAllowedPrincipals`, receberá um erro de falha na Concessão de permissões. Nesse ponto, você deve ver a tabela de destino sendo compartilhada AWS RAM com a conta do consumidor em Permissões, Permissões de dados.

Aceite um compartilhamento de recursos de AWS RAM

> **Note**
>
> Essa etapa é necessária somente para o compartilhamento Conta da AWS baseado, não para o compartilhamento baseado na organização.

1. Faça login no AWS console https://console.aws.amazon.com/connect/usando o administrador do data lake da conta do consumidor usando o nome de usuário do IAM (o padrão é DatalakeAdminConsumer) e a senha especificados durante a criação da CloudFormation pilha.
2. No AWS RAM console, no painel de navegação, em Compartilhado comigo, Compartilhamentos de recursos, escolha o recurso compartilhado do Lake Formation. O status deve ser Pendente.

3. Selecione Ações e Concessão.
4. Confirme os detalhes do recurso e selecione Aceitar compartilhamento de recursos.

   Nesse ponto, o administrador do data lake da conta do consumidor deve ser capaz de encontrar o recurso compartilhado no console do Lake Formation (https://console.aws.amazon.com/lakeformation/) em Catálogo de dados, bancos de dados.

Criar um link de recurso para a tabela compartilhada

- Siga as instruções em Etapa 3: Implementar o compartilhamento entre contas usando o método de controle de acesso baseado em tags (etapa 6) para criar um link de recurso para uma tabela compartilhada. Dê um nome ao link do recurso `amazon_reviews_table_named_resource_resource_link`. Crie o link do recurso no banco de dados `lakeformation_tutorial_cross_account_database_consumer`.

Conceda permissão de dados para a tabela compartilhada ao consumidor

Para conceder permissão de dados para a tabela compartilhada ao consumidor, conclua as etapas a seguir:

1. No console do Lake Formation (https://console.aws.amazon.com/lakeformation/), em Permissões, Permissões do Data lake, escolha Conceder.
2. Para Entidade principais, selecione Usuários e perfis do IAM e selecione o usuário `DataAnalyst`.
3. Em Tags do LF ou recursos de catálogo, selecione Recursos de catálogo de dados nomeados.
4. Em Bancos de dados, selecione o banco de dados `lakeformation_tutorial_cross_account_database_named_resource`. Se você não encontrar o banco de dados na lista suspensa, selecione Carregar mais.
5. Em Tabelas, selecione a tabela `amazon_reviews_table_named_resource`.
6. Para Permissões de tabela e coluna, selecione Selecionar e Descrever em Permissões de tabela.
7. Selecione Conceder.

Conceda permissão de dados para o link de recurso ao consumidor

Além de conceder permissão ao usuário do Data Lake para acessar a tabela compartilhada, você também precisa conceder permissão ao usuário do Data Lake para acessar o link do recurso.

1. No console do Lake Formation (https://console.aws.amazon.com/lakeformation/), em Permissões, Permissões do Data lake, escolha Conceder.
2. Para Entidade principais, selecione Usuários e perfis do IAM e selecione o usuário `DataAnalyst`.
3. Em Tags do LF ou recursos de catálogo, selecione Recursos de catálogo de dados nomeados.
4. Em Bancos de dados, selecione o banco de dados `lakeformation_tutorial_cross_account_database_consumer`. Se você não encontrar o banco de dados na lista suspensa, selecione Carregar mais.
5. Em Tabelas, selecione a tabela `amazon_reviews_table_named_resource_resource_link`.
6. Para Permissões de links de recursos, selecione Descrever em Permissões de links de recursos.
7. Selecione Conceder.

   Nesse ponto, o usuário analista de dados na conta do consumidor deve ser capaz de encontrar o banco de dados e o link do recurso e consultar a tabela compartilhada por meio do console do Athena.

   Caso contrário, confirme se o seguinte está configurado corretamente:

   - O link do recurso é criado para a tabela compartilhada
   - Você concedeu ao usuário acesso à tabela compartilhada pela conta do produtor
   - Você concedeu ao usuário acesso ao link de recurso e ao banco de dados para o qual o link de recurso foi criado

Etapa 5: limpar AWS os recursos

Para evitar cobranças indesejadas Conta da AWS, você pode excluir os AWS recursos usados neste tutorial.

1. Faça login no console do Lake Formation https://console.aws.amazon.com/lakeformation/usando a conta do produtor e exclua ou altere o seguinte:

- AWS Resource Access Manager compartilhamento de recursos
- Tags do Lake Formation
- CloudFormation pilha
- Configurações do Lake Formation
- AWS Glue Data Catalog

2. Faça login no console do Lake Formation https://console.aws.amazon.com/lakeformation/usando a conta do consumidor e exclua ou altere o seguinte:

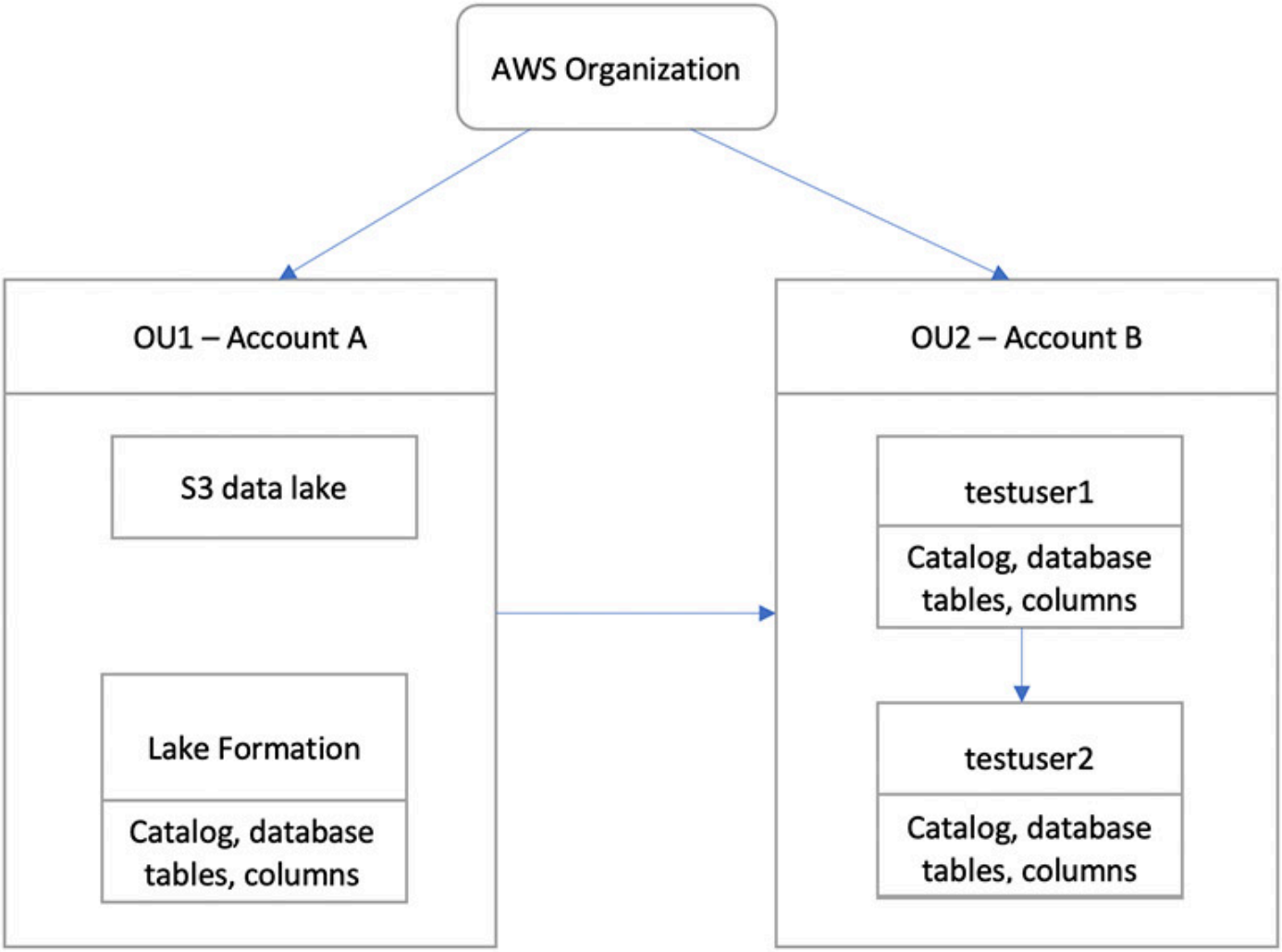
   - Tags do Lake Formation
   - CloudFormation pilha

Como compartilhar um data lake usando o controle de acesso refinado do Lake Formation

Este tutorial fornece step-by-step instruções sobre como você pode compartilhar conjuntos de dados de forma rápida e fácil usando o Lake Formation ao gerenciar vários Contas da AWS com AWS Organizations. Você define permissões específicas para controlar o acesso a dados confidenciais.

Os procedimentos a seguir também mostram como um administrador de data lake da Conta A pode fornecer acesso refinado à Conta B, e como um usuário na Conta B, atuando como administrador de dados, pode conceder acesso refinado à tabela compartilhada para outros usuários em suas contas. Os administradores de dados em cada conta podem delegar o acesso de forma independente a seus próprios usuários, dando autonomia a cada equipe ou linha de negócios (LOB).

O caso de uso pressupõe que você esteja usando AWS Organizations para gerenciar seu Contas da AWS. O usuário da Conta A em uma unidade organizacional (OU1) concede acesso aos usuários da Conta B em OU2. Você pode usar a mesma abordagem quando não estiver usando organizações, como quando você tem apenas algumas contas. O diagrama a seguir ilustra o controle de acesso refinado de conjuntos de dados em um data lake. O data lake está disponível na Conta A. O administrador do data lake da Conta A fornece acesso refinado à Conta B. O diagrama também mostra que um usuário da Conta B fornece acesso em nível de coluna à tabela de data lake da Conta A para outro usuário na Conta B.



Tópicos

- Público-alvo
- Pré-requisitos
- Etapa 1: Forneça acesso refinado a outra conta
- Etapa 2: Forneça acesso refinado a um usuário na mesma conta

Público-alvo

Este tutorial é destinado a administradores de dados, engenheiros de dados e analistas de dados. A tabela a seguir lista os perfis usados neste tutorial:

Perfil	Description
Administrador do IAM	Usuário que tem a política gerenciada pela AWS : `AdministratorAccess` .
Administrador do data lake	Usuário que tem a política AWS `AWSLakeFormationDataAdmin` gerenciada: anexado à função.
Analista de dados	Usuário que tem a política AWS gerenciada: `AmazonAthenaFullAccess` anexado.

Pré-requisitos

Antes de começar este tutorial, você deve ter um Conta da AWS que possa ser usado para entrar como usuário administrativo com as permissões corretas. Para obter mais informações, consulte Inicial completa AWS tarefas de configuração.

O tutorial pressupõe que você esteja familiarizado com o IAM. Para obter informações sobre o IAM, consulte o Guia do usuário do IAM.

Você precisa dos seguintes recursos para este tutorial:

- Duas unidades organizacionais:
  - OU1 — Contém a conta A
  - OU2 — Contém a conta B
- Um local (bucket) do data lake do Amazon S3 na Conta A.
- Um usuário administrador de data lake na Conta A. Você pode criar um administrador de data lake usando o console Lake Formation (https://console.aws.amazon.com/lakeformation/) ou a `PutDataLakeSettings` operação da API Lake Formation.
- Lake Formation configurado na Conta A e o local do data lake do Amazon S3 registrada com o Lake Formation na Conta A.
- Dois usuários na Conta B com as seguintes políticas gerenciadas pelo IAM:
  - testuser1 — tem as políticas AWS `AWSLakeFormationDataAdmin` gerenciadas anexadas.
  - testuser2 — Tem a política AWS `AmazonAthenaFullAccess` gerenciada anexada.
- Um banco de dados testdb no banco de dados Lake Formation para a Conta B.

Etapa 1: Forneça acesso refinado a outra conta

Saiba como um administrador de data lake da Conta A fornece acesso refinado à Conta B.

Conceda acesso refinado a outra conta

1. Faça login Console de gerenciamento da AWS https://console.aws.amazon.com/connect/na Conta A como administrador do data lake.
2. Abra o console do Lake Formation (https://console.aws.amazon.com/lakeformation/) e escolha Começar.
3. No painel de navegação, escolha Bancos de dados.
4. Selecione Criar banco de dados.
5. Na seção informações do Banco de dados, selecione Banco de dados.
6. Em Nome, digite um nome (para este tutorial, usamos `sampledb01`).
7. Certifique-se de que a opção Usar somente o controle de acesso do IAM para novas tabelas nesse banco de dados não esteja selecionada. Deixar essa opção desmarcada nos permite controlar o acesso a partir do Lake Formation.
8. Selecione Criar banco de dados.
9. Na página Bancos de dados, selecione seu banco de dados `sampledb01`.
10. No menu Ações, selecione Conceder.
11. Na seção Conceder permissões, selecione Conta externa.
12. Para Conta da AWS ID ou ID AWS da organização, insira o ID da conta da Conta B em OU2.
13. Em Tabela, selecione a tabela à qual você deseja que a Conta B tenha acesso (neste caso, usamos a tabela `acc_a_area`). Opcionalmente, você pode conceder acesso às colunas dentro da tabela, o que fazemos neste caso.
14. Em Incluir colunas, selecione as colunas às quais você deseja que a Conta B tenha acesso (para este caso, concedemos permissões para digitar, nomear e identificadores).
15. Em Colunas, selecione Incluir colunas.
16. Em Permissões de tabela, selecione Selecionar.
17. Para Permissões concedidas, marque Selecionar. As permissões concedidas são necessárias para que os usuários administradores na Conta B possam conceder permissões a outros usuários na Conta B.
18. Selecione Conceder.

19. No painel de navegação, selecione Tabelas.
20. Você pode ver uma conexão ativa na seção Contas da AWS e AWS organizações com acesso.

Crie um link de recurso

Serviços integrados, como o Amazon Athena, não podem acessar diretamente bancos de dados ou tabelas entre contas. Portanto, você precisa criar um link de recurso para que o Athena possa acessar links de recursos em sua conta para bancos de dados e tabelas em outras contas. Crie um link de recurso para a tabela (`acc_a_area`) para que os usuários da Conta B possam consultar seus dados com o Athena.

1. Faça login no AWS console https://console.aws.amazon.com/connect/na Conta B como`testuser1`.
2. No console do Lake Formation (https://console.aws.amazon.com/lakeformation/), no painel de navegação, escolha Tabelas. Você deve ver as tabelas às quais a Conta A forneceu acesso.
3. Selecione a tabela `acc_a_area`.
4. No menu Ações, selecione Criar link de recurso.
5. Em Nome do link do recurso, digite um nome (para este tutorial, `acc_a_area_rl`).
6. Em Banco de dados, selecione o banco de dados (`testdb`).
7. Escolha Criar.
8. No painel de navegação, selecione Tabelas.
9. Selecione a tabela `acc_b_area_rl`.
10. No menu Ações, selecione Exibir dados.

    Você será redirecionado para o console do Athena, onde deverá ver o banco de dados e a tabela.

    Agora você pode executar uma consulta na tabela para ver o valor da coluna para o qual o acesso foi fornecido ao testuser1 da Conta B.

Etapa 2: Forneça acesso refinado a um usuário na mesma conta

Esta seção mostra como um usuário na Conta B (`testuser1`), atuando como administrador de dados, fornece acesso refinado a outro usuário na mesma conta (`testuser2`) ao nome da coluna na tabela compartilhada `aac_b_area_rl`.

Conceda acesso refinado a um usuário na mesma conta

1. Faça login no AWS console https://console.aws.amazon.com/connect/na Conta B como`testuser1`.
2. No console do Lake Formation, no painel de navegação, selecione Tabelas.

   É possível conceder permissões em uma tabela por meio do link do recurso. Para fazer isso, na página Tabelas, selecione o link do recurso `acc_b_area_rl`, e no menu Ações, selecione Conceder no destino.
3. Na seção Conceder permissões, selecione Minha conta.
4. Para Usuários e perfis do IAM, selecione o usuário `testuser2`.
5. Em Coluna, escolha o nome da coluna.
6. Em Permissões de tabela, selecione Selecionar.
7. Selecione Conceder.

   Após criar um link de recurso, somente você poderá visualizá-lo e acessá-lo. Para permitir que outros usuários da sua conta acessem o link do recurso, você precisa conceder permissões no próprio link do recurso. Você precisa conceder as permissões Descrever ou Descartar. Na página Tabelas, selecione sua tabela novamente e, no menu Ações, selecione Conceder.
8. Na seção Conceder permissões, selecione Minha conta.
9. Para Usuários e perfis do IAM, selecione o usuário `testuser2`.
10. Para permissões de links de recursos, selecione Descrever.
11. Selecione Conceder.
12. Faça login no AWS console na Conta B como`testuser2`.

    No console do Athena (https://console.aws.amazon.com/athena/), você deve ver o banco de dados e a tabela. `acc_b_area_rl` Agora você pode executar uma consulta na tabela para ver o valor da coluna que `testuser2` tem acesso.

Permissões de integração ao Lake Formation

AWS Lake Formation usa o AWS Glue Data Catalog (Catálogo de dados) para armazenar metadados para os data lakes do Amazon S3 e fontes de dados externas, como o Amazon Redshift, na forma de catálogos, bancos de dados e tabelas. Os metadados no Data Catalog são organizados em uma hierarquia de dados de três níveis que inclui catálogos, bancos de dados e tabelas. Ele organiza dados de várias fontes em contêineres lógicos chamados catálogos. Bancos de dados são coleções de tabelas. O catálogo de dados também contém links de recursos, que são links para bancos de dados e tabelas compartilhados em contas externas e são usados para acesso entre contas aos dados no data lake. Cada AWS conta tem um catálogo de dados por AWS região.

O Lake Formation fornece um modelo de permissões do sistema de gerenciamento de banco de dados relacional (RDBMS) para conceder ou revogar o acesso a catálogos, bancos de dados, tabelas e colunas no Data Catalog com dados subjacentes no Amazon S3.

Antes de aprender sobre os detalhes do modelo de permissões do Lake Formation, é útil revisar as seguintes informações básicas:

- Data lakes gerenciados pelo Lake Formation residem em locais designados no Amazon Simple Storage Service (Amazon S3). O Data Catalog também contém objetos de catálogo. Cada catálogo representa dados de fontes, como data warehouses do Amazon Redshift, bancos de dados do Amazon DynamoDB e fontes de dados de terceiros, como Snowflake, MySQL, e mais de trinta fontes de dados externas, que são integradas por meio de conectores federados.
- O Lake Formation mantém um catálogo de dados que contém metadados sobre dados de origem a serem importados para seus data lakes, como dados em logs e bancos de dados relacionais, e sobre dados em seus data lakes no Amazon S3. O Data Catalog também contém metadados sobre dados de fontes de dados externas que não sejam o Amazon S3. Os metadados são organizados como catálogos, bancos de dados e tabelas. As tabelas de metadados contêm esquema, localização, particionamento e outras informações sobre os dados que elas representam. Bancos de dados de metadados são coleções de tabelas.
- O catálogo de dados do Lake Formation é o mesmo catálogo de dados usado pelo AWS Glue. Você pode usar crawlers do AWS Glue para criar tabelas do catálogo de dados e pode usar tarefas de extração, transformação e carregamento (ETL) do AWS Glue para preencher os dados subjacentes em seus data lakes.
- Os catálogos, bancos de dados e tabelas no Data Catalog são chamados de recursos do Data Catalog. As tabelas no catálogo de dados são chamadas de tabelas de metadados para

diferenciá-las das tabelas nas fontes de dados ou dos dados tabulares no Amazon S3. Os dados para os quais as tabelas de metadados apontam no Amazon S3 ou nas fontes de dados são chamados de dados subjacentes.

- Um principal é um usuário ou função, um usuário ou grupo do Amazon Quick, um usuário ou grupo que se autentica no Lake Formation por meio de um provedor de SAML ou, para controle de acesso entre contas, um ID da AWS conta, ID da organização ou ID da unidade organizacional.
- AWS Glueos rastreadores criam tabelas de metadados, mas você também pode criar tabelas de metadados manualmente com o console do Lake Formation, a API ou o (). AWS Command Line Interface AWS CLI Ao criar uma tabela de metadados, você deve especificar uma localização. Quando você cria um banco de dados, o local é opcional. Os locais das tabelas podem ser locais do Amazon S3 ou locais de fonte de dados, como um banco de dados do Amazon Relational Database Service (Amazon RDS). Os locais do banco de dados são sempre locais do Amazon S3.
- Serviços que se integram ao Lake Formation, como Amazon Athena e Amazon Redshift, podem acessar o catálogo de dados para obter metadados e verificar a autorização para executar consultas. Para obter uma lista completa de serviços integrados, consulte AWS integrações de serviços com Lake Formation.

Tópicos

- Visão geral das permissões do Lake Formation
- Referência de personas e permissões do IAM do Lake Formation
- Alterando as configurações padrão do seu data lake
- Permissões implícitas do Lake Formation
- Referência de permissões do Lake Formation
- Integrar o Centro de Identidade do IAM
- Adicionar uma localização do Amazon S3 ao seu data lake
- Modo de acesso híbrido
- Criação de objetos no AWS Glue Data Catalog
- Importação de dados usando fluxos de trabalho no Lake Formation

Visão geral das permissões do Lake Formation

Há dois tipos principais de permissões no AWS Lake Formation:

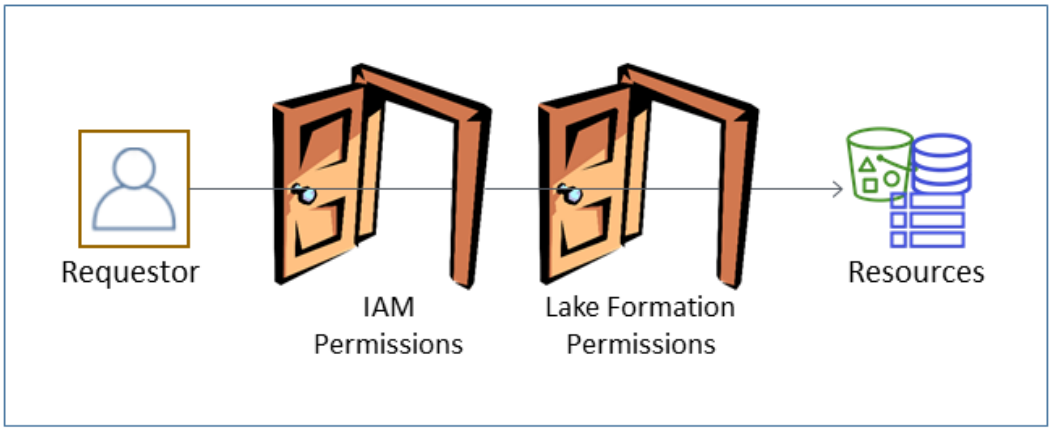
- Acesso aos metadados – Permissões nos recursos do catálogo de dados (Permissões do catálogo de dados).

  Com essas permissões, as entidades principais podem criar, ler, atualizar e excluir bancos de dados e tabelas de metadados no catálogo de dados.

- Acesso aos dados subjacentes — Permissões em locais no Amazon Simple Storage Service (Amazon S3) (permissões de acesso a dados e permissões de localização de dados).
  - As permissões do data lake possibilitam que as entidades principais leiam e gravem dados em locais subjacentes do Amazon S3 – dados apontados pelos recursos do catálogo de dados.
  - Com as permissões de localização de dados, as entidades principais podem criar e alterar bancos de dados e tabelas de metadados que apontam para os locais específicos do Amazon S3.

Para ambas as áreas, o Lake Formation usa uma combinação de permissões e permissões do Lake Formation AWS Identity and Access Management (IAM). O modelo de permissões do IAM consiste em políticas do IAM. O modelo de permissões do Lake Formation é implementado como DBMS-style GRANT/REVOKE comandos, como`Grant SELECT on` *`tableName`* `to` *`userName`*.

Quando uma entidade principal faz uma solicitação para acessar os recursos do catálogo de dados ou os dados subjacentes, para que a solicitação seja bem-sucedida, ela deve passar pelas verificações de permissão do IAM e do Lake Formation.



As permissões do Lake Formation controlam o acesso aos recursos do catálogo de dados, aos locais do Amazon S3 e aos dados subjacentes nesses locais. As permissões do IAM controlam o acesso ao Lake Formation, APIs AWS Glue e recursos. Portanto, embora você possa ter a permissão do Lake Formation para criar uma tabela de metadados no catálogo de dados (`CREATE_TABLE`), sua

operação falhará se você não tiver a permissão do IAM na API `glue:CreateTable`. (Por que uma permissão do `glue:`? Porque o Lake Formation usa o catálogo de dados do AWS Glue.)

> **Note**
>
> As permissões do Lake Formation se aplicam somente na região em que foram concedidas.

AWS Lake Formation exige que cada diretor (usuário ou função) seja autorizado a realizar ações nos recursos gerenciados pelo Lake Formation. Uma entidade principal recebe as autorizações necessárias do administrador do data lake ou de outra entidade principal com as permissões para conceder as permissões do Lake Formation.

Ao conceder uma permissão de Lake Formation a uma entidade principal, você pode, opcionalmente, conceder a capacidade de passar essa permissão para outra entidade principal.

Você pode usar a API do Lake Formation, a AWS Command Line Interface (AWS CLI) ou as páginas Permissões de dados e Localizações de dados do console do Lake Formation para conceder e revogar as permissões do Lake Formation.

Métodos para controle de acesso de alta granularidade

Com um data lake, o objetivo é ter um controle de acesso de alta granularidade aos dados. No Lake Formation, isso significa controle de acesso de alta granularidade aos recursos do catálogo de dados e aos locais do Amazon S3. Você pode obter um controle de acesso de alta granularidade com um dos seguintes métodos.

Método	Permissões do Lake Formation	Permissões do IAM	Comentários
Método 1	Abra o	Fine-grained	Esse é o método padrão para compatibilidade com versões anteriores com o AWS Glue. • Aberto significa que a permissão especial `Super` é concedida ao grupo `IAMAllowedPrincipals` , onde `IAMAllowedPrincipals` é criada

Método	Permissões do Lake Formation	Permissões do IAM	Comentários
			automaticamente e inclui todos os usuários e perfis do IAM que têm permissão para acessar seus recursos do catálogo de dados por meio de suas políticas do IAM, e a permissão `Super` permite que uma entidade principal execute todas as operações suportadas do Lake Formation no banco de dados ou na tabela em que foi concedida. Isso efetivamente faz com que o acesso aos recursos do catálogo de dados e aos locais do Amazon S3 seja controlado exclusivamente pelas políticas do IAM. Para obter mais informações, consulte Alterando as configurações padrão do seu data lake e Atualizando as permissões AWS Glue de dados para o modelo AWS Lake Formation. • Fine-grainedsignifica que as políticas do IAM controlam todo o acesso aos recursos do catálogo de dados e aos buckets individuais do Amazon S3. No console do Lake Formation, esse método aparece como Use apenas controle de acesso do IAM.

Método	Permissões do Lake Formation	Permissões do IAM	Comentários
Método 2	Fine-grained	Coarse-grained	Este é o método recomendado. • Fine-grainedacesso significa conceder permissões limitadas do Lake Formation a diretores individuais sobre os recursos do Catálogo de Dados, os locais do Amazon S3 e os dados subjacentes nesses locais. • Coarse-grainedsignifica permissões mais amplas em operações individuais e no acesso aos locais do Amazon S3. Por exemplo, uma política de IAM de baixa granularidade pode incluir `"glue:*"` ou `"glue:Create*"` ao invés de `"glue:CreateTables"` , deixando que as permissões do Lake Formation controlem se uma entidade principal pode ou não criar objetos de catálogo. Isso também significa dar às entidades principais acesso às APIs de que precisam para realizar seu trabalho, mas bloquear outras APIs e recursos. Por exemplo, você pode criar uma política do IAM que permita que uma entidade principal crie recursos do catálogo de dados e crie e execute fluxos de trabalho, mas não permita a criação de conexões do AWS Glue ou funções definidas pelo usuário. Veja os exemplos mais adiante nesta seção.

> **Important**
>
> Esteja ciente do seguinte:
>
> - Por padrão, o Lake Formation tem as configurações de Controle de acesso apenas para uso do IAM habilitadas para compatibilidade com o comportamento existente do catálogo de dados do AWS Glue. Recomendamos que você desative essas configurações após a transição para o uso das permissões do Lake Formation. Para obter mais informações, consulte Alterando as configurações padrão do seu data lake.
> - Os administradores do Data Lake e os criadores de bancos de dados têm permissões implícitas do Lake Formation que você precisa entender. Para obter mais informações, consulte Permissões implícitas do Lake Formation.

Controle de acesso a metadados

Para controle de acesso aos recursos do catálogo de dados, a discussão a seguir pressupõe controle de acesso de alta granularidade com permissões do Lake Formation e controle de acesso de baixa granularidade com políticas do IAM.

Há dois métodos distintos para conceder permissões do Lake Formation nos recursos do catálogo de dados:

- Controle de acesso a recursos nomeados – Com esse método, você concede permissões em bancos de dados ou tabelas específicos especificando nomes de bancos de dados ou tabelas. As concessões têm o seguinte formato:

  Conceda Permissões a entidades principais sobre recursos [com opção de concessão].

  Com a opção de concessão, você pode permitir que o beneficiário conceda as permissões a outras entidades principais.
- Tag-based controle de acesso — Com esse método, você atribui um ou mais LF-Tags bancos de dados, tabelas e colunas do Catálogo de Dados e concede permissões em um ou mais LF-Tags aos principais. Cada um LF-Tag é um par de valores-chave, como. `department=sales` Um diretor que tenha LF-Tags essa correspondência LF-Tags em um recurso do Catálogo de Dados pode acessar esse recurso. Esse método é recomendado para data lakes com um grande número de bancos de dados e tabelas. Isso é explicado em detalhes em Controle de acesso baseado em tags do Lake Formation.

As permissões que uma entidade principal tem sobre um recurso são a união das permissões concedidas pelos dois métodos.

A tabela a seguir resume as permissões disponíveis do Lake Formation nos recursos do catálogo de dados. Os títulos das colunas indicam o recurso no qual a permissão é concedida.

Catálogo	Banco de dados	Tabela
CREATE_DATABASE	CREATE_TABLE	ALTER
	ALTER	DROP
	DROP	DESCRIBE
	DESCRIBE	SELECT*
		INSERT*
		DELETE*

Por exemplo, a permissão `CREATE_TABLE` é concedida em um banco de dados. Isso significa que a entidade principal tem permissão para criar tabelas nesse banco de dados.

As permissões com um asterisco (*) são concedidas nos recursos do catálogo de dados, mas se aplicam aos dados subjacentes. Por exemplo, a permissão `DROP` em uma tabela de metadados permite que você remova a tabela do catálogo de dados. No entanto, a permissão `DELETE` concedida na mesma tabela permite que você exclua os dados subjacentes da tabela no Amazon S3, usando, por exemplo, uma instrução `DELETE` do SQL. Com essas permissões, você também pode visualizar a tabela no console do Lake Formation e recuperar informações sobre a tabela com a API do AWS Glue. Portanto, `SELECT`, `INSERT` e `DELETE` são permissões do catálogo de dados e permissões de acesso aos dados.

Ao conceder `SELECT` em uma tabela, você pode adicionar um filtro que inclua ou exclua uma ou mais colunas. Isso permite um controle de acesso de alta granularidade nas colunas da tabela de metadados, limitando as colunas que os usuários de serviços integrados podem ver ao executar consultas. Esse recurso não está disponível usando apenas as políticas do IAM.

Também há uma permissão especial chamada `Super`. A permissão `Super` permite que uma entidade principal execute todas as operações suportadas do Lake Formation no banco de dados

ou na tabela em que ela foi concedida. Essa permissão pode coexistir com as outras permissões do Lake Formation. Por exemplo, você pode conceder `Super`, `SELECT` e `INSERT` em uma tabela de metadados. A entidade principal pode realizar todas as ações suportadas na tabela e, quando você revoga `Super`, as permissões `SELECT` e `INSERT` permanecem.

Para obter detalhes sobre cada permissão, consulte Referência de permissões do Lake Formation.

> **Important**
>
> Para poder ver uma tabela do catálogo de dados criada por outro usuário, você deve ter pelo menos uma permissão do Lake Formation na tabela. Se você tiver pelo menos uma permissão na tabela, também poderá ver o banco de dados que contém a tabela.

Você pode conceder ou revogar as permissões do catálogo de dados usando o console do Lake Formation, a API ou o AWS Command Line Interface (AWS CLI). Veja a seguir um exemplo de um AWS CLI comando que concede ao usuário `datalake_user1` permissão para criar tabelas no `retail` banco de dados.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "CREATE_TABLE" --resource '{ "Database": {"Name":"retail"}}'
```

Veja a seguir um exemplo de uma política do IAM de controle de acesso de baixa granularidade que complementa o controle de acesso de alta granularidade com as permissões do Lake Formation. Ele permite todas as operações em qualquer banco de dados ou tabela de metadados.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "glue:*Database*",
                "glue:*Table*",
                "glue:*Partition*"
            ],
```

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

O próximo exemplo também é de baixa granularidade, mas um pouco mais restritivo. Ele permite operações somente para leitura em todos os bancos de dados e tabelas de metadados no catálogo de dados na conta e região designadas.

JSON

```
{
    "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:*"
        }
    ]
}
```

Compare essas políticas com a política a seguir, que implementa um controle de IAM-based acesso refinado. Ele concede permissões somente em um subconjunto de tabelas no banco de dados de metadados de gerenciamento de relacionamento com o cliente (CRM) na conta e região designadas.

JSON

```
{
    "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*"
            ]
        }
    ]
}
```

Para obter mais exemplos de políticas de controle de acesso de baixa granularidade, consulte Referência de personas e permissões do IAM do Lake Formation.

Controle de acesso a dados subjacente

Quando um AWS serviço integrado solicita acesso aos dados em um local do Amazon S3 que é controlado pelo acesso, o Lake AWS Lake Formation Formation fornece credenciais temporárias para acessar os dados.

Para permitir que o Lake Formation controle o acesso aos dados subjacentes em um local do Amazon S3, você registra esse local no Lake Formation.

Depois de registrar uma localização no Amazon S3, você pode começar a conceder as seguintes permissões do Lake Formation:

- Permissões de acesso aos dados (`SELECT`, `INSERT` e `DELETE`) nas tabelas do catálogo de dados que apontam para esse local.
- Permissões de localização de dados nesse local.

As permissões de localização de dados do Lake Formation controlam a capacidade de criar recursos do catálogo de dados que apontam para locais específicos do Amazon S3. As permissões de localização de dados fornecem uma camada extra de segurança aos locais dentro do data lake. Ao conceder a permissão `CREATE_TABLE` ou `ALTER` a uma entidade principal, você também

concede permissões de localização de dados para limitar os locais para os quais a entidade principal pode criar ou alterar tabelas de metadados.

Os locais do Amazon S3 são buckets ou prefixos em um bucket, mas não objetos individuais do Amazon S3.

Você pode conceder permissões de localização de dados a uma entidade principal usando o console do Lake Formation, a API ou a AWS CLI. A forma geral de uma subvenção é a seguinte:

```
grant DATA_LOCATION_ACCESS to principal on S3 location [with grant option]
```

Se você incluir `with grant option`, o beneficiário poderá conceder as permissões a outras entidades principais.

Lembre-se de que as permissões do Lake Formation sempre funcionam em combinação com as permissões AWS Identity and Access Management (IAM) para um controle de acesso refinado. Para read/write permissões sobre dados subjacentes do Amazon S3, as permissões do IAM são concedidas da seguinte forma:

Ao registrar um local, você especifica uma função do IAM que concede read/write permissões nesse local. A Lake Formation assume essa função ao fornecer credenciais temporárias para serviços integrados. AWS Uma função típica pode ter a seguinte política anexada, em que o local registrado é o bucket `awsexamplebucket`.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "s3:PutObject",
                "s3:GetObject",
                "s3:DeleteObject"
            ],
            "Resource": [
                "arn:aws:s3:::amzn-s3-demo-bucket/*"
            ]
        },
```

```
        {
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::amzn-s3-demo-bucket"
            ]
        }
    ]
}
```

O Lake Formation fornece uma função vinculada ao serviço que você pode usar durante o registro para criar automaticamente políticas como essa. Para obter mais informações, consulte Usar perfis vinculados ao serviço para o Lake Formation.

Portanto, registrar um local do Amazon S3 concede as permissões necessárias `s3:` do IAM nesse local, onde as permissões são especificadas pela função usada para registrar o local.

> **Important**
>
> Evite registrar um bucket do Amazon S3 que tenha o Solicitante paga ativado. Para buckets registrados no Lake Formation, a função usada para registrar o bucket é sempre vista como solicitante. Se o bucket for acessado por outra AWS conta, o proprietário do bucket será cobrado pelo acesso aos dados se a função pertencer à mesma conta do proprietário do bucket.

Para read/write acessar os dados subjacentes, além das permissões do Lake Formation, os diretores também precisam da permissão do `lakeformation:GetDataAccess` IAM. Com essa permissão, o Lake Formation concede a solicitação de credenciais temporárias para acessar os dados.

JSON

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

```
            "Action": "lakeformation:GetDataAccess",
            "Resource": "*"
        }
    ]
}
```

Na política acima, você deve definir o parâmetro Resource como '*' (all). A especificação de qualquer outro recurso para essa permissão não é aceita. Essa configuração garante que o Lake Formation possa gerenciar o acesso aos dados em todo o seu ambiente de data lake de forma eficiente.

> **Note**
>
> O Amazon Athena exige que o usuário tenha a permissão `lakeformation:GetDataAccess`. Outros serviços integrados exigem que sua função de execução subjacente tenha a permissão `lakeformation:GetDataAccess`.

Essa permissão está incluída nas políticas sugeridas no Referência de personas e permissões do IAM do Lake Formation.

Resumindo, para permitir que as entidades principais do Lake Formation leiam e gravem dados subjacentes com acesso controlado pelas permissões do Lake Formation:

- Registre os locais do Amazon S3 que contêm os dados com o Lake Formation.
- As entidades principais que criam tabelas do catálogo de dados que apontam para locais de dados subjacentes devem ter permissões de localização de dados.
- As entidades principais que leem e gravam dados subjacentes devem ter permissões de acesso aos dados do Lake Formation nas tabelas do catálogo de dados que apontam para os locais de dados subjacentes.
- As entidades principais que leem e gravam dados subjacentes devem ter a permissão `lakeformation:GetDataAccess` do IAM quando o local dos dados subjacentes é registrado no Lake Formation.

> **Note**
>
> O modelo de permissões do Lake Formation não impede o acesso aos locais do Amazon S3 por meio da API ou do console do Amazon S3 se você tiver acesso a eles por meio

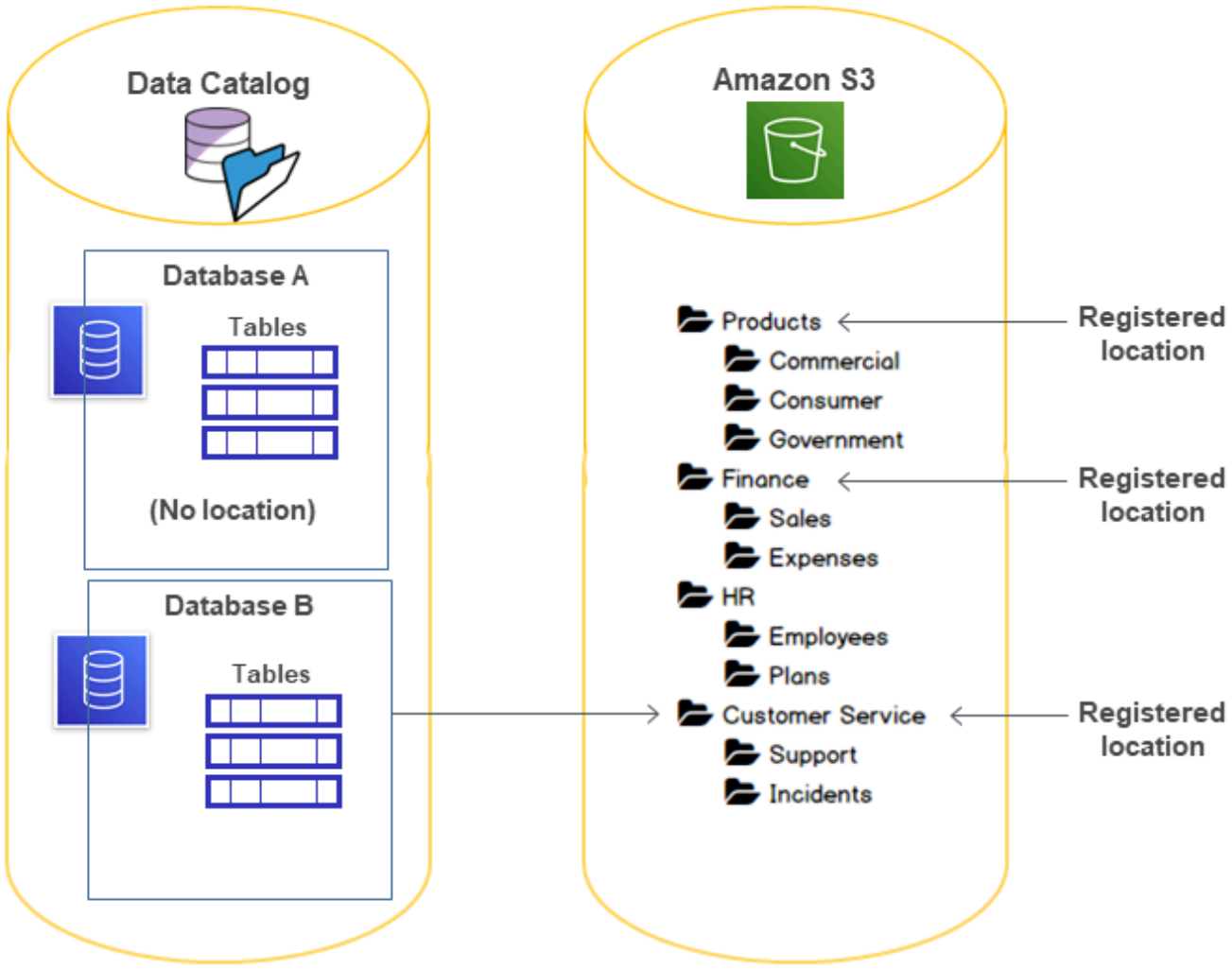
de políticas do IAM ou do Amazon S3. Você pode anexar políticas do IAM às entidades principais para bloquear esse acesso.

Saiba mais sobre permissões de localização de dados

As permissões de localização de dados governam o resultado das operações de criação e atualização nos bancos de dados e tabelas do catálogo de dados. As regras são as seguintes:

- Uma entidade principal deve ter permissões de localização de dados explícitas ou implícitas em um local do Amazon S3 para criar ou atualizar um banco de dados ou tabela que especifique esse local.
- A permissão explícita `DATA_LOCATION_ACCESS` é concedida usando o console, a API ou AWS CLI.
- As permissões implícitas são concedidas quando um banco de dados tem uma propriedade de localização que aponta para um local registrado, a entidade principal tem a permissão `CREATE_TABLE` no banco de dados e a entidade principal tenta criar uma tabela nesse local ou em um local secundário.
- Se uma entidade principal receber permissões de localização de dados em um local, a entidade principal terá permissões de localização de dados em todos os locais secundários.
- Um diretor não precisa de permissões de localização de dados para realizar read/write operações nos dados subjacentes. É suficiente ter as permissões de acesso aos dados `SELECT` ou `INSERT`. As permissões de localização de dados se aplicam somente à criação de recursos do catálogo de dados que apontam para o local.
- As operações de partição (`CreatePartitionBatchCreatePartition`,,`UpdatePartition`) não exigem permissões de localização de dados quando a localização da partição é filha da localização da tabela já registrada no Lake Formation.

Considere o cenário mostrado no diagrama a seguir.



Neste diagrama:

- Os buckets do Amazon S3 `Products`, `Finance` e `Customer Service` estão registrados no Lake Formation.
- `Database A` não tem propriedade de localização e `Database B` tem uma propriedade de localização que aponta para o bucket `Customer Service`.
- O usuário `datalake_user` tem `CREATE_TABLE` nos dois bancos de dados.
- O usuário `datalake_user` recebeu permissões de localização de dados somente no bucket `Products`.

A seguir estão os resultados quando o usuário `datalake_user` tenta criar uma tabela de catálogo em um banco de dados específico em um determinado local.

Local onde **`datalake_user`** tenta criar uma tabela

Banco de dados e localização	Sucesso ou falha	Motivo
Banco de dados A em `Finance/Sales`	Falha	Sem permissão de localização de dados
Banco de dados A em `Products`	Sucesso	Possui permissão de localização de dados
Banco de dados A em `HR/Plans`	Sucesso	O local não está registrado
Banco de dados B em `Customer Service/Incidents`	Sucesso	O banco de dados tem propriedade de localização em `Customer Service`

Para saber mais, consulte:

- Adicionar uma localização do Amazon S3 ao seu data lake
- Referência de permissões do Lake Formation
- Referência de personas e permissões do IAM do Lake Formation

Referência de personas e permissões do IAM do Lake Formation

Esta seção lista algumas personas sugeridas de Lake Formation e suas permissões do AWS Identity and Access Management (IAM) sugeridas. Para obter informações sobre as permissões do Lake Formation, consulte the section called “Referência de permissões do Lake Formation”.

AWS Lake Formation personas

A tabela a seguir lista as AWS Lake Formation personas sugeridas.

Personas do Lake Formation

Pessoa	Description
Administrador do IAM (superusuário)	(Obrigatório) Usuário que pode criar usuários e perfis do IAM. Tem a política `AdministratorAccess` AWS gerenciada. Tem todas as permissões em todos os recursos do Lake Formation. Pode adicionar administradores de data lake. Não é possível conceder permissões do Lake Formation se também não for designado como administrador do data lake.
Administrador do data lake	(Obrigatório) Usuário que pode registrar locais do Amazon S3, acessar o catálogo de dados, criar bancos de dados, criar e executar fluxos de trabalho, conceder permissões do Lake Formation a outros usuários e visualizar registros. AWS CloudTrail Tem menos permissões do IAM do que o administrador do IAM, mas o suficiente para administrar o data lake. Não é possível adicionar outros administradores de data lake.
Administrador somente para leitura	(Opcional) Usuário que pode visualizar as entidades principais, os recursos, as permissões e os logs AWS CloudTrail do catálogo de dados, sem a permissão para fazer atualizações.
Engenheiro de dados	(Opcional) Usuário que pode criar bancos de dados, criar e executar crawlers e fluxos de trabalho e conceder permissões do Lake Formation nas tabelas do catálogo de dados que os crawlers e fluxos de trabalho criam. Recomendamos tornar todos os engenheiros de dados criadores de bancos de dados. Para obter mais informações, consulte Criação de um banco de dados.
Analista de dados	(Opcional) Usuário que pode executar consultas no data lake usando, por exemplo, Amazon Athena. Tem permissões suficientes apenas para executar consultas.
Função do fluxo de trabalho	(Obrigatório) Função que executa um fluxo de trabalho em nome de um usuário. Você especifica esse perfil ao criar um fluxo de trabalho a partir de um esquema.

> **Note**
>
> No Lake Formation, os administradores de data lake adicionados após a criação do banco de dados podem conceder permissões, mas não têm automaticamente permissões de acesso aos dados, como SELECT ou DESCRIBE. Os administradores que criam bancos de dados recebem permissões `SUPER` nesses bancos de dados. Esse comportamento é intencional. Embora todos os administradores possam conceder a si mesmos as permissões necessárias, essas permissões não são aplicadas automaticamente aos recursos preexistentes. Portanto, os administradores devem conceder explicitamente a si mesmos acesso aos bancos de dados que existiam antes de receberem privilégios de administrador.

AWS políticas gerenciadas para Lake Formation

Você pode conceder as permissões AWS Identity and Access Management (IAM) necessárias para trabalhar AWS Lake Formation usando políticas AWS gerenciadas e políticas em linha. As seguintes políticas AWS gerenciadas estão disponíveis para Lake Formation.

AWS política gerenciada: AWSLake FormationDataAdmin

AWSLakeFormationDataAdmina política concede acesso administrativo AWS Lake Formation e serviços relacionados, como AWS Glue o gerenciamento de lagos de dados.

Você pode anexar `AWSLakeFormationDataAdmin` aos seus usuários, grupos e funções.

Detalhes da permissão

- `CloudTrail`— Permite que os diretores visualizem os AWS CloudTrail registros. Isso é necessário para analisar quaisquer erros na configuração do data lake.
- `Glue` – Permite que as entidades principais visualizem, criem e atualizem tabelas de metadados e bancos de dados no catálogo de dados. Isso inclui operações de API que começam com `Get`, `List`, `Create`, `Update`, `Delete` e `Search`. Isso é necessário para gerenciar os metadados das tabelas do data lake.
- `IAM` – Permite que as entidades principais recuperem informações sobre usuários, perfis e políticas do IAM vinculadas às funções. Isso é necessário para que o administrador de dados revise e liste os usuários e perfis do IAM para conceder permissões ao Lake Formation.

- `Lake Formation` – Concede aos administradores do data lake as permissões necessárias do Lake Formation para gerenciar os data lakes.
- `S3` – Permite que as entidades principais recuperem informações sobre os buckets do Amazon S3 e suas localizações para configurar a localização dos dados para os data lakes.

```
"Statement": [
        {
            "Sid": "AWSLakeFormationDataAdminAllow",
            "Effect": "Allow",
            "Action": [
                "lakeformation:*",
                "cloudtrail:DescribeTrails",
                "cloudtrail:LookupEvents",
                "glue:CreateCatalog",
  "glue:UpdateCatalog",
                "glue:DeleteCatalog",
  "glue:GetCatalog",
         "glue:GetCatalogs",
                "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**
>
> A política `AWSLakeFormationDataAdmin` não concede todas as permissões necessárias para administradores de data lake. Permissões adicionais são necessárias para criar e executar fluxos de trabalho e registrar locais com a função `AWSServiceRoleForLakeFormationDataAccess` vinculada ao serviço. Para obter mais informações, consulte Crie um administrador de data lake e Usar perfis vinculados ao serviço para o Lake Formation.

AWS política gerenciada: AWSLake FormationCrossAccountManager

AWSLakeFormationCrossAccountManagera política fornece acesso entre contas a AWS Glue recursos por meio do Lake Formation e concede acesso de leitura a outros serviços necessários, como AWS Organizations AWS RAM e.

Você pode anexar `AWSLakeFormationCrossAccountManager` aos seus usuários, grupos e funções.

Detalhes da permissão

Esta política inclui as seguintes permissões.

- `Glue` – Permite que as entidades principais definam ou excluam a política de recursos do catálogo de dados para controle de acesso.
- `Organizations` – Permite que as entidades principais recuperem as informações da conta e da unidade organizacional (OU) de uma organização.
- `ram:CreateResourceShare` – Permite que as entidades principais criem um compartilhamento de recursos.
- `ram:UpdateResourceShare` – Permite que as entidades principais modifiquem algumas propriedades do compartilhamento de recursos especificado.
- `ram:DeleteResourceShare` – Permite que as entidades principais excluam o compartilhamento de recursos especificado.
- `ram:AssociateResourceShare` – Permite que as entidades principais adicionem a lista especificada de entidades principais e a lista de recursos a um compartilhamento de recursos.
- `ram:DisassociateResourceShare` – Permite que as entidades principais removam as entidades principais ou recursos especificados da participação no compartilhamento de recursos especificado.
- `ram:GetResourceShares` – Permite que as entidades principais recuperem detalhes sobre os compartilhamentos de recursos que você possui ou que são compartilhados com você.
- `ram:RequestedResourceType` – Permite que as entidades principais recuperem o tipo de recurso (banco de dados, tabela ou catálogo).
- `AssociateResourceSharePermission`— Permite que os diretores adicionem ou substituam a AWS RAM permissão de um tipo de recurso incluído em um compartilhamento de recursos. Você pode ter exatamente uma permissão associada a cada tipo de recurso no compartilhamento de recursos.

JSON

```
{
    "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": {
                "ArnLike": {
                    "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 política gerenciada: AWSGlue ConsoleFullAccess

AWSGlueConsoleFullAccessa política concede acesso total aos AWS Glue recursos quando uma identidade à qual a política está anexada usa Console de gerenciamento da AWS o. Se você seguir a convenção de nomenclatura para os recursos especificados nesta política, os usuários poderão acessar todos os recursos do console. Essa política geralmente é anexada aos usuários do AWS Glue console.

Além disso, AWS Glue a Lake Formation assume a função de serviço `AWSGlueServiceRole` para permitir o acesso a serviços relacionados, incluindo Amazon Elastic Compute Cloud (Amazon EC2), Amazon Simple Storage Service (Amazon S3) e Amazon. CloudWatch

AWS managed policy:LakeFormationDataAccessServiceRolePolicy

Essa política é anexada a um perfil vinculado ao serviço chamado `ServiceRoleForLakeFormationDataAccess`, o qual possibilita que o serviço execute ações em recursos quando solicitado. Não é possível anexar essa política às suas identidades do IAM.

Essa política permite que os AWS serviços integrados do Lake Formation, como Amazon Athena o Amazon Redshift, usem a função vinculada ao serviço para descobrir os recursos do Amazon S3.

Para obter mais informações, consulte, Usar perfis vinculados ao serviço para o Lake Formation.

Detalhes de permissões

Esta política inclui a seguinte permissão.

- `s3:ListAllMyBuckets`: exibe uma lista de todos os buckets de propriedade do remetente autenticado da solicitação.

JSON

```
{
 "Version":"2012-10-17",
 "Statement": [
  {
   "Sid": "LakeFormationDataAccessServiceRolePolicy",
   "Effect": "Allow",
   "Action": [
    "s3:ListAllMyBuckets"
   ],
   "Resource": [
    "arn:aws:s3:::*"
   ]
  }
 ]
}
```

Atualizações do Lake Formation nas políticas AWS gerenciadas

Veja detalhes sobre as atualizações das políticas AWS gerenciadas do Lake Formation desde que esse serviço começou a rastrear essas mudanças.

Alteração	Descrição	Data
Política `AWSLakeFormationCrossAccountManager` atualizada do Lake Formation.	O Lake Formation aprimorou a AWSLakeFormationCrossAccountManagerpolítica substituindo o operador de `StringLike` condição pelo `ArnLike` operador que permite ao IAM realizar a verificação do formato ARN.	Janeiro de 2025
Política `AWSLakeFormationDataAdmin` atualizada do Lake Formation.	A Lake Formation aprimorou a AWSLakeFormationDataAdminpolítica adicionando o seguinte AWS Glue Data Catalog CRUD APIs como parte do recurso de vários catálogos. • cola: CreateCatalog • cola: UpdateCatalog • cola: DeleteCatalog • cola: GetCatalog • cola: GetCatalogs Essa alteração de política gerenciada é para garantir que a pessoa de administrador do Lake Formation, por padrão, tenha permissão do IAM para essas novas operações.	Dezembro de 2024
Política `AWSLakeFormationCrossAccountManager` atualizada do Lake Formation.	A Lake Formation aprimorou a AWSLakeFormationCrossAccountManagerpolítica adicionando elementos de Sid à declaração de política.	Março de 2024

Alteração	Descrição	Data
Política `AWSLakeFormationDataAdmin` atualizada do Lake Formation.	A Lake Formation aprimorou a AWSLakeFormationDataAdminpolítica adicionando um elemento Sid à declaração de política e removendo uma ação redundante.	Março de 2024
Política `LakeFormationDataAccessServiceRolePolicy` atualizada do Lake Formation.	A Lake Formation aprimorou a LakeFormationDataAccessServiceRolePolicypolítica adicionando um elemento Sid à declaração de política.	Fevereiro de 2024
Política `AWSLakeFormationCrossAccountManager` atualizada do Lake Formation.	A Lake Formation aprimorou a AWSLakeFormationCrossAccountManagerpolítica adicionando uma nova permissão para permitir o compartilhamento de dados entre contas no modo de acesso híbrido.	Outubro de 2023
Política `AWSLakeFormationCrossAccountManager` atualizada do Lake Formation.	A Lake Formation aprimorou a AWSLakeFormationCrossAccountManagerpolítica para criar apenas um compartilhamento de recursos por conta de destinatário quando o recurso é compartilhado pela primeira vez. Todos os recursos compartilhados posteriormente com a mesma conta são vinculados ao mesmo compartilhamento de recursos.	6 de maio de 2022
O Lake Formation passou a monitorar as alterações.	A Lake Formation começou a monitorar as mudanças em suas políticas AWS gerenciadas.	6 de maio de 2022

Permissões sugeridas por personas

A seguir estão as permissões sugeridas para cada persona. O administrador do IAM não está incluído porque esse usuário tem todas as permissões em todos os recursos.

Tópicos

- Permissões de administrador do data lake
- Permissões de administrador somente para leitura
- Permissões de engenheiro de dados
- Permissões de analista de dados
- Permissões da função de fluxo de trabalho

Permissões de administrador do data lake

> **Important**
>
> Nas políticas a seguir, *<account-id>* substitua por um número de AWS conta válido e *<workflow_role>* substitua pelo nome de uma função que tenha permissões para executar um fluxo de trabalho, conforme definido emPermissões da função de fluxo de trabalho.

Tipo de política	Política
AWS políticas gerenciadas	• `AWSLakeFormationDataAdmin` • `LakeFormationDataAccessServiceRolePolicy` (política de perfil vinculado ao serviço) • `AWSGlueConsoleFullAccess` (Opcional) • `CloudWatchLogsReadOnlyAccess` (Opcional) • `AWSLakeFormationCrossAccountManager` (Opcional) • `AmazonAthenaFullAccess` (Opcional)

Tipo de política	Política
	Para obter informações sobre as políticas AWS gerenciadas opcionais, consultethe section called “Crie um administrador de data lake”.

Política embutida (para criar a função vinculada ao serviço Lake Formation)	{
	"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"
	}
	]
	}

Tipo de política	Política

(Opcional) Política embutida (política de senha para a função de fluxo de trabalho). Isso é necessário somente se o administrador do data lake criar e executar fluxos de trabalho.	{
	"Version":"2012-10-17",
	"Statement": [
	{
	"Sid": "PassRolePermissions",
	"Effect": "Allow",
	"Action": [
	"iam:PassRole"
	],
	"Resource": [
	"arn:aws:iam:: *111122223333* :role/*<workflow_role>* "
	]
	}
	]
	}

(Opcional) Política embutida (se sua conta estiver concedendo ou recebendo permissões entre contas do Lake Formation). Essa política serve para aceitar ou rejeitar convites de compartilhamento de AWS RAM recursos e para permitir a concessão de permissões entre contas às organizações. `ram:EnableSharingWithAwsOrganization` é necessário somente para administradores de data lake na conta de AWS Organizations gerenciamento.	{
	"Version":"2012-10-17",
	"Statement": [
	{
	"Effect": "Allow",
	"Action": [
	"ram:AcceptResourceShareInvitation",
	"ram:RejectResourceShareInvitation",
	"ec2:DescribeAvailabilityZones",
	"ram:EnableSharingWithAwsOrganization"
	],
	"Resource": "*"
	}
	]
	}

Permissões de administrador somente para leitura

Tipo de política	Política
Política embutida (básica)	{ "Version":"2012-10-17", "Statement":[{ "Effect":"Allow", "Action":["lakeformation:GetEffectivePermissionsForPath", "lakeformation:ListPermissions", "lakeformation:ListDataCellsFilter", "lakeformation:GetDataCellsFilter", "lakeformation:SearchDatabasesByLFTags", "lakeformation:SearchTablesByLFTags", "lakeformation:GetLFTag", "lakeformation:ListLFTags", "lakeformation:GetResourceLFTags", "lakeformation:ListLakeFormationOptions", "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",

Tipo de política	Política

	"s3:ListAllMyBuckets",
	"s3:GetBucketAcl",
	"iam:ListUsers",
	"iam:ListRoles",
	"iam:GetRole",
	"iam:GetRolePolicy"
	],
	"Resource":"*"
	},
	{
	"Effect":"Deny",
	"Action":[
	"lakeformation:PutDataLakeSettings"
	],
	"Resource":"*"
	}
	]
	}

Permissões de engenheiro de dados

> **Important**
>
> Nas políticas a seguir, *<account-id>* substitua por um número de AWS conta válido e *<workflow_role>* substitua pelo nome da função do fluxo de trabalho.

Tipo de política	Política
AWS política gerenciada	`AWSGlueConsoleFullAccess`
Política embutida (básica)	{ "Version":"2012-10-17", "Statement": [{ "Effect": "Allow", "Action": [

Tipo de política	Política
	"lakeformation:GetDataAccess", "lakeformation:GrantPermissions", "lakeformation:RevokePermissions", "lakeformation:BatchGrantPermissions", "lakeformation:BatchRevokePermissions", "lakeformation:ListPermissions", "lakeformation:AddLFTagsToResource", "lakeformation:RemoveLFTagsFromResource", "lakeformation:GetResourceLFTags", "lakeformation:ListLFTags", "lakeformation:GetLFTag", "lakeformation:SearchTablesByLFTags", "lakeformation:SearchDatabasesByLFTags", "lakeformation:GetWorkUnits", "lakeformation:GetWorkUnitResults", "lakeformation:StartQueryPlanning", "lakeformation:GetQueryState", "lakeformation:GetQueryStatistics"], "Resource": "*" }] }

Tipo de política	Política

Política embutida (para operações em tabelas controladas, incluindo operações dentro de transações)	{
	"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": "*"
	}
	]
	}

Tipo de política	Política

Política embutida (para controle de acesso a metadados usando o método de controle de acesso baseado em tags (LF-TBAC) do Lake Formation)	{
	"Version":"2012-10-17",
	"Statement": [
	{
	"Effect": "Allow",
	"Action": [
	"lakeformation:AddLFTagsToResource",
	"lakeformation:RemoveLFTagsFromResou
	rce",
	"lakeformation:GetResourceLFTags",
	"lakeformation:ListLFTags",
	"lakeformation:GetLFTag",
	"lakeformation:SearchTablesByLFTags"
	,
	"lakeformation:SearchDatabasesByLFTa
	gs"
	],
	"Resource": "*"
	}
	]
	}

Política embutida (política de senha para a função de fluxo de trabalho)	{
	"Version":"2012-10-17",
	"Statement": [
	{
	"Sid": "PassRolePermissions",
	"Effect": "Allow",
	"Action": [
	"iam:PassRole"
	],
	"Resource": [
	"arn:aws:iam:: *111122223
	333* :role/*<workflow_role>* "
	]
	}
	]
	}

Permissões de analista de dados

Tipo de política	Política
AWS política gerenciada	`AmazonAthenaFullAccess`
Política embutida (básica)	`{ "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": "*" }] }`
(Opcional) Política embutida (para operações em tabelas controladas, incluindo operações dentro de transações)	`{ "Version":"2012-10-17", "Statement": [ { "Effect": "Allow", "Action": [ "lakeformation:StartTransaction", "lakeformation:CommitTransaction", "lakeformation:CancelTransaction", "lakeformation:ExtendTransaction",`

Tipo de política	Política
	"lakeformation:DescribeTransaction", "lakeformation:ListTransactions", "lakeformation:GetTableObjects", "lakeformation:UpdateTableObjects", "lakeformation:DeleteObjectsOnCancel "], "Resource": "*" }] }

Permissões da função de fluxo de trabalho

Essa função tem as permissões necessárias para executar um fluxo de trabalho. Você especifica uma função com essas permissões ao criar um fluxo de trabalho.

> **Important**
>
> Nas políticas a seguir, *<region>* substitua por um identificador de AWS região válido (por exemplous-east-1), *<account-id>* por um número de AWS conta válido, *<workflow_role>* pelo nome da função do fluxo de trabalho e *<your-s3-cloudtrail-bucket>* pelo caminho do Amazon S3 para seus AWS CloudTrail registros.

Tipo de política	Política
AWS política gerenciada	AWSGlueServiceRole
Política embutida (acesso a dados)	{ "Version": "2012-10-17", "Statement": [{ "Sid": "Lakeformation", "Effect": "Allow", "Action": ["lakeformation:GetDataAccess",

Tipo de política	Política

	<pre>"lakeformation:GrantPermissions"
	],
	"Resource": "*"
	}
	]
	}</pre>

Política embutida (política de senha para a função de fluxo de trabalho)	<pre>{
	"Version":"2012-10-17",
	"Statement": [
	{
	"Sid": "PassRolePermissions",
	"Effect": "Allow",
	"Action": [
	"iam:PassRole"
	],
	"Resource": [
	"arn:aws:iam:: 111122223333 :role/<workflow_role> "
	]
	}
	]
	}</pre>

Política embutida (para ingerir dados fora do data lake, por exemplo, AWS CloudTrail registros)	<pre>{
	"Version":"2012-10-17",
	"Statement": [
	{
	"Effect": "Allow",
	"Action": ["s3:GetObject", "s3:ListBucket"],
	"Resource": ["arn:aws:s3::: <your-s3-cloudtrail-bucket> /*"]
	}
	]
	}</pre>

Alterando as configurações padrão do seu data lake

Para manter a compatibilidade com versões anterioresAWS Glue, AWS Lake Formation tem as seguintes configurações iniciais de segurança:

- A permissão `Super` é concedida ao grupo `IAMAllowedPrincipals` em todos os recursos existentes do catálogo de dados do AWS Glue.
- As configurações “Usar somente o controle de acesso ao IAM” estão habilitadas para novos recursos do catálogo de dados.

Essas configurações efetivamente fazem com que o acesso aos recursos do catálogo de dados e aos locais do Amazon S3 seja controlado exclusivamente por políticas AWS Identity and Access Management (IAM). As permissões individuais do Lake Formation não estão em vigor.

O grupo `IAMAllowedPrincipals` inclui todos os usuários e perfis do IAM que possuem permissão para acessar os recursos do seu catálogo de dados por meio de suas políticas do IAM. A permissão `Super` possibilita que uma entidade principal execute todas as operações suportadas do Lake Formation no banco de dados ou na tabela em que ela foi concedida.

Para alterar as configurações de segurança para que o acesso aos recursos do catálogo de dados (bancos de dados e tabelas) seja gerenciado pelas permissões do Lake Formation, faça o seguinte:

1. Altere as configurações de segurança padrão para novos recursos. Para instruções, consulte Alterar o modelo de permissão padrão ou usar o modo de acesso híbrido.
2. Altere as configurações dos recursos existentes do catálogo de dados. Para instruções, consulte Atualizando as permissões AWS Glue de dados para o modelo AWS Lake Formation.

Alterando as configurações de segurança padrão usando a operação da API do Lake Formation **`PutDataLakeSettings`**

Você também pode alterar as configurações de segurança padrão usando a operação da PutDataLakeSettingsAPI Lake Formation. Essa ação usa como argumentos um ID de catálogo opcional e uma DataLakeSettingsestrutura.

Para impor metadados e controle de acesso aos dados subjacentes pelo Lake Formation em novos bancos de dados e tabelas, codifique a estrutura `DataLakeSettings` da seguinte forma.

> **Note**
>
> *<AccountID>*Substitua por um ID de AWS conta válido e *<Username>* por um nome de usuário do IAM válido. Você pode especificar mais de um usuário como administrador de data lake.

```
{
    "DataLakeSettings": {
        "DataLakeAdmins": [
            {
                "DataLakePrincipalIdentifier":
 "arn:aws:iam::<AccountId>:user/<Username>"
            }
        ],
        "CreateDatabaseDefaultPermissions": [],
        "CreateTableDefaultPermissions": []
    }
}
```

Você também pode codificar a estrutura da seguinte maneira. Omitir o parâmetro `CreateDatabaseDefaultPermissions` ou `CreateTableDefaultPermissions` é equivalente a passar uma lista vazia.

```
{
    "DataLakeSettings": {
        "DataLakeAdmins": [
            {
                "DataLakePrincipalIdentifier":
 "arn:aws:iam::<AccountId>:user/<Username>"
            }
        ]
    }
}
```

Essa ação revoga efetivamente todas as permissões do grupo `IAMAllowedPrincipals` Lake Formation em novos bancos de dados e tabelas. Ao criar um banco de dados, você pode substituir essa configuração.

Para aplicar metadados e controle de acesso aos dados subjacentes somente pelo IAM em novos bancos de dados e tabelas, codifique a estrutura `DataLakeSettings` da seguinte forma.

```
{
    "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"
                ]
            }
        ]
    }
}
```

Isso concede ao Lake Formation `Super` permissão para o grupo `IAMAllowedPrincipals` em novos bancos de dados e tabelas. Ao criar um banco de dados, você pode substituir essa configuração.

> **Note**
>
> Na estrutura `DataLakeSettings` anterior, o único valor permitido para `DataLakePrincipalIdentifier` é `IAM_ALLOWED_PRINCIPALS` e o único valor permitido para `Permissions` é `ALL`.

Permissões implícitas do Lake Formation

AWS Lake Formation concede as seguintes permissões implícitas aos administradores do data lake, criadores de banco de dados e criadores de tabelas.

Administradores de data lake

- Tenha acesso `Describe` a todos os recursos do catálogo de dados, exceto aos recursos compartilhados de outra conta diretamente com uma entidade principal diferente. Esse acesso não pode ser revogado por um administrador.
- Tenha permissões de localização de dados em todos os lugares no data lake.
- Pode conceder ou revogar o acesso a quaisquer recursos no catálogo de dados a qualquer entidade principal (inclusive a si mesmo). Esse acesso não pode ser revogado por um administrador.
- Pode criar bancos de dados no catálogo de dados.
- Pode conceder permissão para criar um banco de dados para outro usuário.

> **Note**
>
> Os administradores do data lake podem registrar locais do Amazon S3 somente se tiverem permissões do IAM para fazer isso. As políticas de administrador de data lake sugeridas neste guia concedem essas permissões. Além disso, os administradores do data lake não têm permissões implícitas para eliminar bancos de dados ou alter/drop tabelas criados por outras pessoas. No entanto, eles podem conceder a si mesmos permissões para fazer isso.

Para obter mais informações sobre administradores de data lake, consulte Crie um administrador de data lake.

Criadores de catálogos

- Tenha todas as permissões de catálogo nos catálogos que eles criam, tenham permissões nos bancos de dados e tabelas que eles criam no catálogo e possam conceder a outros diretores da mesma AWS conta permissão para criar bancos de dados e tabelas no catálogo. Um criador de catálogo que também tenha a política `AWSLakeFormationCrossAccountManager` AWS gerenciada pode conceder permissões no catálogo a outras AWS contas ou organizações.

  Os administradores do data lake podem usar o console ou a API do Lake Formation para designar criadores de catálogos.

  > **Note**
  >
  > Os criadores de catálogos não têm permissões implícitas em bancos de dados e tabelas que outras pessoas criam no catálogo.

  Para acessar mais informações sobre como criar catálogos, consulte Trazendo seus dados para o AWS Glue Data Catalog.

Criadores de banco de dados

- Tenha todas as permissões de banco de dados nos bancos de dados que eles criam, tenham permissões nas tabelas que eles criam no banco de dados e possam conceder a outros diretores da mesma AWS conta permissão para criar tabelas no banco de dados. Um criador de banco de dados que também tenha a política `AWSLakeFormationCrossAccountManager` AWS gerenciada pode conceder permissões no banco de dados para outras AWS contas ou organizações.

  Os administradores do data lake podem usar o console ou a API do Lake Formation para designar criadores de banco de dados.

  > **Note**
  >
  > Os criadores de banco de dados não possui permissões implícitas em tabelas que outras pessoas criam no banco de dados.

  Para obter mais informações, consulte Criação de um banco de dados.

Criadores de tabelas

- Tenha todas as permissões nas tabelas que eles criam.

- Podem conceder permissões em todas as tabelas que eles criam para entidades principais na mesma conta da AWS .
- Podem conceder permissões em todas as tabelas que eles criam para outras AWS contas ou organizações se tiverem a política `AWSLakeFormationCrossAccountManager` AWS gerenciada.
- Pode visualizar os bancos de dados que contêm as tabelas que eles criam.

Referência de permissões do Lake Formation

Para realizar AWS Lake Formation operações, os diretores precisam tanto das permissões do Lake Formation quanto das permissões AWS Identity and Access Management (IAM). Normalmente, você concede permissões do IAM usando políticas de controle de acesso de baixa granularidade, conforme descrito em the section called “Visão geral das permissões do Lake Formation”. Você pode conceder permissões do Lake Formation usando o console, a API ou o AWS Command Line Interface (AWS CLI).

Para saber como conceder ou revogar permissões do Lake Formation, consulte the section called “Conceder permissões de dados” e the section called “Conceder permissões de localização de dados”.

> **Note**
>
> Os exemplos nesta seção mostram como conceder permissões às entidades principais na mesma conta da AWS . Para obter exemplos de concessões entre contas, consulte the section called “Cross-account compartilhamento de dados”.

Permissões do Lake Formation por tipo de recurso

A seguir estão as permissões válidas do Lake Formation disponíveis para cada tipo de recurso:

Recurso	Permissão
`Catalog`	`ALL` (`Super`), superusuário
	`ALTER`

Recurso	Permissão
	CREATE_DATABASE
	DESCRIBE
	DROP
Database	ALL (Super)
	ALTER
	CREATE_TABLE
	DESCRIBE
	DROP
Table	ALL (Super)
	ALTER
	DELETE
	DESCRIBE
	DROP
	INSERT
	SELECT
View	ALL (Super)
	SELECT
	DESCRIBE
	DROP
Data Catalog	CREATE_DATABASE

Recurso	Permissão
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
	DROP
	INSERT
	SELECT
Resource link - Database or Table	DESCRIBE
	DROP

Recurso	Permissão
Table with data filters	DESCRIBE
	DROP
	SELECT
Table with column filter	SELECT

Tópicos

- Concessão e revogação da Lake Formation AWS CLI Comandos da do
- Permissões do Lake Formation

Concessão e revogação da Lake Formation AWS CLI Comandos da do

Cada descrição de permissão nesta seção inclui exemplos de como conceder a permissão usando um AWS CLI comando. A seguir estão as sinopses da Lake Formation e dos comandos. grant-permissions revoke-permissions AWS CLI

```
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>]
```

Para obter descrições detalhadas desses comandos, consulte grant-permissions e revoke-permissions na Referência de comandos da AWS CLI . Esta seção fornece informações adicionais sobre a opção `--principal`.

O valor da opção `--principal` é um dos seguintes:

- Nome de recurso da Amazon (ARN) para um usuário ou função AWS Identity and Access Management (IAM)
- ARN para um usuário ou grupo que se autentica por meio de um provedor SAML, como o Microsoft Active Directory Federation Service (AD FS)
- ARN para um usuário ou grupo do Amazon Quick
- Para permissões entre contas, uma ID AWS da conta, uma ID da organização ou uma ID da unidade organizacional
- Para usuário ou grupo do Centro de Identidade do IAM, ARN do usuário ou grupo do Centro de Identidade do IAM.

Veja a seguir a sintaxe e exemplos para todos os tipos `--principal`.

Entidade principal é um usuário do IAM

Sintaxe:

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/<user-name>
```

Exemplo:

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1
```

Entidade principal é um perfil do IAM

Sintaxe:

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:role/<role-name>
```

Exemplo:

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:role/workflowrole
```

Entidade principal é um usuário que se autentica por meio de um provedor SAML

Sintaxe:

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:saml-
provider/<SAMLproviderName>:user/<user-name>
```

Exemplos:

```
--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
```

Entidade principal é um grupo que se autentica por meio de um provedor SAML

Sintaxe:

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:saml-
provider/<SAMLproviderName>:group/<group-name>
```

Exemplos:

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:saml-provider/
idp1:group/data-scientists
```

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:saml-provider/
AthenaLakeFormationOkta:group/my-group
```

Principal é usuário do Amazon Quick Enterprise Edition

Sintaxe:

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:<region>:<account-
id>:user/<namespace>/<user-name>
```

> **Note**
>
> Para *<namespace>*, você deve especificar `default`.

Exemplo:

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:us-
east-1:111122223333:user/default/bi_user1
```

Principal é um grupo Amazon Quick Enterprise Edition

Sintaxe:

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:<region>:<account-
id>:group/<namespace>/<group-name>
```

> **Note**
>
> Para *<namespace>*, você deve especificar `default`.

Exemplo:

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:us-
east-1:111122223333:group/default/data_scientists
```

Principal é uma AWS conta

Sintaxe:

```
--principal DataLakePrincipalIdentifier=<account-id>
```

Exemplo:

```
--principal DataLakePrincipalIdentifier=111122223333
```

Entidade principal é uma organização

Sintaxe:

```
--principal DataLakePrincipalIdentifier=arn:aws:organizations::<account-
id>:organization/<organization-id>
```

Exemplo:

```
--principal
 DataLakePrincipalIdentifier=arn:aws:organizations::111122223333:organization/o-
abcdefghijkl
```

Entidade principal é uma unidade organizacional

Sintaxe:

```
--principal DataLakePrincipalIdentifier=arn:aws:organizations::<account-
id>:ou/<organization-id>/<organizational-unit-id>
```

Exemplo:

```
--principal DataLakePrincipalIdentifier=arn:aws:organizations::111122223333:ou/o-
abcdefghijkl/ou-ab00-cdefghij
```

A entidade principal é um usuário ou grupo de identidade do Centro de Identidade do IAM

Example:User

```
--principal DataLakePrincipalIdentifier=arn:aws:identitystore:::user/<UserID>
```

Example:Group:

```
--principal DataLakePrincipalIdentifier=arn:aws:identitystore:::group/<GroupID>
```

A entidade principal é um grupo do IAM (**`IAMAllowedPrincipals`**)

Por padrão, o Lake Formation define a permissão `Super` em todos os bancos de dados e tabelas no Catálogo de Dados para um grupo chamado `IAMAllowedPrincipals`. Se essa permissão de grupo existir em um banco de dados ou tabela, todas as entidades principais da sua conta terão acesso ao recurso por meio das políticas de entidade principal do IAM para o AWS Glue. A compatibilidade com versões anteriores será fornecida quando você começar a usar as permissões do Lake Formation para proteger os recursos do Catálogo de Dados que antes eram protegidos pelas políticas do IAM para o AWS Glue.

Ao usar o Lake Formation para gerenciar permissões para seus recursos do Catálogo de Dados, primeiro você precisa revogar a permissão `IAMAllowedPrincipals` dos recursos ou optar pelo

modo de acesso híbrido às entidades principais e aos recursos para que as permissões do Lake Formation funcionem.

Exemplo:

```
--principal DataLakePrincipalIdentifier=IAM_Allowed_Principals
```

A entidade principal é um grupo do IAM (**`ALLIAMPrincipals`**)

Quando você concede permissões a um grupo de `ALLIAMPrincipals` em um recurso do Catálogo de Dados, cada entidade principal da conta tem acesso ao recurso do Catálogo de Dados usando as permissões do Lake Formation e do IAM.

Exemplo:

```
--principal DataLakePrincipalIdentifier=123456789012:IAMPrincipals
```

Permissões do Lake Formation

Esta seção contém as permissões disponíveis do Lake Formation que você pode conceder às entidades principais.

`ALTER`

Permissão	Concedido neste recurso	O beneficiário também precisa
`ALTER`	`DATABASE`	`glue:UpdateDatabase`
`ALTER`	`TABLE`	`glue:UpdateTable`
`ALTER`	`LF-Tag`	`lakeformation:UpdateLFTag`

Uma entidade principal com essa permissão pode alterar os metadados de um banco de dados ou tabela no catálogo de dados. Para tabelas, você pode alterar o esquema da coluna e adicionar parâmetros da coluna. Você não pode alterar as colunas nos dados subjacentes para os quais uma tabela de metadados aponta.

Se a propriedade que está sendo alterada for um local registrado do Amazon Simple Storage Service (Amazon S3), a entidade principal deverá ter permissões de localização de dados no novo local.

Example

O exemplo a seguir concede a `ALTER` permissão ao usuário `datalake_user1` no banco de dados `retail` na AWS conta 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "ALTER" --resource '{ "Database": {"Name":"retail"}}'
```

Example

O exemplo a seguir concede `ALTER` ao usuário `datalake_user1` na tabela `inventory` no banco de dados `retail`.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "ALTER" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

CRIAR_BANCO DE DADOS

Permissão	Concedido neste recurso	O beneficiário também precisa
`CREATE_DATABASE`	catálogo de dados	`glue:CreateDatabase`

Uma entidade principal com essa permissão pode criar um banco de dados de metadados ou um link de recurso no catálogo de dados. A entidade principal também pode criar tabelas no banco de dados.

Example

O exemplo a seguir concede `CREATE_DATABASE` ao usuário `datalake_user1` na AWS conta 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "CREATE_DATABASE" --resource '{ "Catalog": {}}'
```

Quando uma entidade principal cria um banco de dados no catálogo de dados, nenhuma permissão para os dados subjacentes é concedida. As seguintes permissões adicionais de metadados são concedidas (junto com a capacidade de conceder essas permissões a outras pessoas):

- `CREATE_TABLE` no banco de dados
- Banco de dados da `ALTER`
- Banco de dados da `DROP`

Ao criar um banco de dados, a entidade principal pode opcionalmente especificar um local do Amazon S3. Dependendo se a entidade principal tem permissões de localização de dados, a permissão `CREATE_DATABASE` pode não ser suficiente para criar bancos de dados em todos os casos. É importante ter em mente os três casos a seguir.

Crie um caso de uso de banco de dados	Permissões necessárias
A propriedade do local não é especificada.	`CREATE_DATABASE` é suficiente.
A propriedade de localização é especificada e a localização não é gerenciada pelo Lake Formation (não está registrada).	`CREATE_DATABASE` é suficiente.
A propriedade de localização é especificada e a localização é gerenciada pelo Lake Formation (está registrada).	`CREATE_DATABASE` é obrigatório, além de permissões de localização de dados no local especificado.

`CRIAR_TABELA`

Permissão	Concedido neste recurso	O beneficiário também precisa
`CREATE_TABLE`	`DATABASE`	`glue:CreateTable`

Uma entidade principal com essa permissão pode criar uma tabela de metadados ou um link de recurso no catálogo de dados dentro do banco de dados especificado.

Example

O exemplo a seguir concede ao usuário `datalake_user1` permissão para criar tabelas no `retail` banco de dados na AWS conta 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "CREATE_TABLE" --resource '{ "Database": {"Name":"retail"}}'
```

Quando uma entidade principal cria uma tabela no catálogo de dados, todas as permissões do Lake Formation na tabela são concedidas à entidade principal, com a capacidade de conceder essas permissões a outras pessoas.

Cross-account Subsídios

Se uma conta do proprietário do banco de dados conceder `CREATE_TABLE` a uma conta do destinatário e um usuário na conta do destinatário criar com êxito uma tabela no banco de dados da conta do proprietário, as seguintes regras se aplicam:

- O usuário e os administradores do data lake na conta do destinatário têm todas as permissões do Lake Formation disponíveis. Eles podem conceder permissões na tabela a outras entidades principais de suas contas. Eles não podem conceder permissões às entidades principais na conta do proprietário ou em qualquer outra conta.
- Os administradores do data lake na conta do proprietário podem conceder permissões na tabela a outras entidades principais da conta.

Permissões de localização de dados

Quando você tenta criar uma tabela que aponta para um local do Amazon S3, dependendo se você tem permissões de localização de dados, a permissão `CREATE_TABLE` pode não ser suficiente para criar uma tabela. É importante ter em mente os três casos a seguir.

Crie um caso de uso de tabela	Permissões necessárias
O local especificado não é gerenciado pelo Lake Formation (não está registrado).	`CREATE_TABLE` é suficiente.

Crie um caso de uso de tabela	Permissões necessárias
O local especificado é gerenciado pelo Lake Formation (está registrado), e o banco de dados que o contém não tem propriedade de localização ou tem uma propriedade de localização que não seja um prefixo do Amazon S3 da localização da tabela.	`CREATE_TABLE` é obrigatório, além de permissões de localização de dados no local especificado.
O local especificado é gerenciado pelo Lake Formation (está registrado), e o banco de dados que o contém tem uma propriedade de localização que aponta para um local registrado e é um prefixo Amazon S3 da localização da tabela.	`CREATE_TABLE` é suficiente.

ACESSO À LOCALIZAÇÃO DE DADOS

Permissão	Concedido neste recurso	O beneficiário também precisa
`DATA_LOCATION_ACCESS`	Local do Amazon S3	(Permissões do Amazon S3 no local, que devem ser especificadas pela função usada para registrar o local.)

Essa é a única permissão de localização de dados. Uma entidade principal com essa permissão pode criar um banco de dados ou tabela de metadados que aponte para a localização especificada do Amazon S3. O local deve ser registrado. Uma entidade principal que tem permissões de localização de dados em um local também tem permissões de localização em locais secundários.

Example

O exemplo a seguir concede permissões de localização de dados de `s3://products/retail` ao usuário `datalake_user1` na conta AWS 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` não é necessário consultar ou atualizar dados subjacentes. Essa permissão se aplica somente à criação de recursos do catálogo de dados.

Para obter mais informações sobre permissões de local de dados, consulte Underlying data access control.

DELETE

Permissão	Concedido neste recurso	O beneficiário também precisa
`DELETE`	`TABLE`	(Nenhuma permissão adicional do IAM é necessária se o local estiver registrado.)

Uma entidade principal com essa permissão pode inserir, atualizar e ler dados subjacentes no local do Amazon S3 especificado pela tabela. A entidade principal também pode visualizar a tabela no console do Lake Formation e recuperar informações sobre a tabela com a API do AWS Glue.

Example

O exemplo a seguir concede a `DELETE` permissão ao usuário na tabela `datalake_user1` no banco de dados `inventory retail` na AWS conta 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DELETE" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

Essa permissão se aplica somente aos dados no Amazon S3 e não aos dados de outros armazenamentos de dados, como o Amazon Relational Database Service (Amazon RDS).

DESCRIBE

Permissão	Concedido neste recurso	O beneficiário também precisa
`DESCRIBE`	Link de recurso de tabela	`glue:GetTable`

Permissão	Concedido neste recurso	O beneficiário também precisa
	Link de recurso de banco de dados	`glue:GetDatabase`
`DESCRIBE`	`DATABASE`	`glue:GetDatabase`
`DESCRIBE`	`TABLE`	`glue:GetTable`

`DESCRIBE`	`LF-Tag`	`glue:GetTable`
		`glue:GetDatabase`
		`lakeformation:GetResourceLFTags`
		`lakeformation:ListLFTags`
		`lakeformation:GetLFTag`
		`lakeformation:SearchTablesByLFTags`
		`lakeformation:SearchDatabasesByLFTags`

Uma entidade principal com essa permissão pode visualizar o banco de dados, a tabela ou o link do recurso especificado. Nenhuma outra permissão do catálogo de dados é concedida implicitamente e nenhuma permissão de acesso aos dados é concedida implicitamente. Bancos de dados e tabelas aparecem nos editores de consultas dos serviços integrados, mas nenhuma consulta pode ser feita neles, a menos que outras permissões do Lake Formation (por exemplo, `SELECT`) sejam concedidas.

Por exemplo, um usuário que tem `DESCRIBE` em um banco de dados pode ver o banco de dados e todos os metadados do banco de dados (descrição, localização e assim por diante). No entanto, o usuário não consegue descobrir quais tabelas o banco de dados contém e não pode descartar, alterar ou criar tabelas no banco de dados. Da mesma forma, um usuário que tem `DESCRIBE` em

uma tabela pode ver a tabela e os metadados da tabela (descrição, esquema, localização e assim por diante), mas não pode descartar, alterar ou executar consultas na tabela.

A seguir estão algumas regras adicionais para `DESCRIBE`:

- Se um usuário tiver outras permissões do Lake Formation em um banco de dados, tabela ou link de recurso, `DESCRIBE` será concedida implicitamente.
- Se um usuário tiver `SELECT` em apenas um subconjunto de colunas para uma tabela (parcial `SELECT`), o usuário estará restrito a ver apenas essas colunas.
- Você não pode conceder `DESCRIBE` a um usuário que tenha seleção parcial em uma tabela. Por outro lado, você não pode especificar listas de inclusão ou exclusão de colunas para tabelas concedidas a `DESCRIBE`.

Example

O exemplo a seguir concede a `DESCRIBE` permissão ao usuário `datalake_user1` no link do recurso de tabela no banco de dados `inventory-link` `retail` na AWS conta 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

Permissão	Concedido neste recurso	O beneficiário também precisa
`DROP`	`DATABASE`	`glue:DeleteDatabase`
`DROP`	`TABLE`	`glue:DeleteTable`
`DROP`	`LF-Tag`	`lakeformation:DeleteLFTag`

`DROP`	Link de recurso de banco de dados
	
	Link de recurso de tabela
		`glue:DeleteDatabase`
		
		`glue:DeleteTable`

Uma entidade principal com essa permissão pode colocar um link de banco de dados, tabela ou recurso no catálogo de dados. Você não pode conceder DROP em um banco de dados a uma conta ou organização externa.

> **Warning**
>
> Eliminar um banco de dados elimina todas as tabelas no banco de dados.

Example

O exemplo a seguir concede a DROP permissão ao usuário no banco de dados `datalake_user1` `retail` na AWS conta 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "DROP" --resource '{ "Database": {"Name":"retail"}}'
```

Example

O exemplo a seguir concede DROP ao usuário `datalake_user1` na tabela `inventory` do banco de dados `retail`.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DROP" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

Example

O exemplo a seguir concede DROP ao usuário `datalake_user1` na tabela o link do recurso `inventory-link` no banco de dados `retail`.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "DROP" --resource '{ "Table": {"DatabaseName":"retail", "Name":"inventory-
link"}}'
```

INSERT

Permissão	Concedido neste recurso	O beneficiário também precisa
INSERT	TABLE	(Nenhuma permissão adicional do IAM é necessária se o local estiver registrado.)

Uma entidade principal com essa permissão pode inserir, atualizar e ler dados subjacentes no local do Amazon S3 especificado pela tabela. A entidade principal também pode visualizar a tabela no console do Lake Formation e recuperar informações sobre a tabela com a API do AWS Glue.

Example

O exemplo a seguir concede a INSERT permissão ao usuário na tabela datalake_user1 no banco de dados inventory retail na AWS conta 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "INSERT" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

Essa permissão se aplica somente aos dados no Amazon S3 e não aos dados em outros armazenamentos de dados, como o Amazon RDS.

SELECT

Permissão	Concedido neste recurso	O beneficiário também precisa
SELECT	• TABLE	(Nenhuma permissão adicional do IAM é necessária se o local estiver registrado.)

Uma entidade principal com essa permissão pode visualizar uma tabela no catálogo de dados e consultar os dados subjacentes no Amazon S3 no local especificado pela tabela. A entidade principal pode visualizar a tabela no console do Lake Formation e recuperar informações sobre a tabela com a API do AWS Glue. Se a filtragem de colunas foi aplicada quando essa permissão foi concedida, a

entidade principal poderá visualizar os metadados somente das colunas incluídas e poderá consultar dados somente das colunas incluídas.

> **Note**
>
> É responsabilidade do serviço de análise integrada aplicar a filtragem de colunas ao processar uma consulta.

Example

O exemplo a seguir concede a `SELECT` permissão ao usuário na tabela `datalake_user1` no banco de dados `inventory retail` na AWS conta 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "SELECT" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

Essa permissão se aplica somente aos dados no Amazon S3 e não aos dados em outros armazenamentos de dados, como o Amazon RDS.

Você pode filtrar (restringir o acesso a) colunas específicas com uma lista de inclusão opcional ou uma lista de exclusão. Uma lista de inclusão especifica as colunas que podem ser acessadas. Uma lista de exclusão especifica as colunas que não podem ser acessadas. Na ausência de uma lista de inclusão ou exclusão, todas as colunas da tabela estão acessíveis.

Os resultados de `glue:GetTable` retornam somente as colunas que o autor da chamada tem permissão para visualizar. Serviços integrados, como Amazon Athena e Amazon Redshift, honram as listas de inclusão e exclusão de colunas.

Example

O exemplo a seguir concede `SELECT` ao usuário `datalake_user1` na tabela `inventory` usando uma lista de inclusão.

```
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

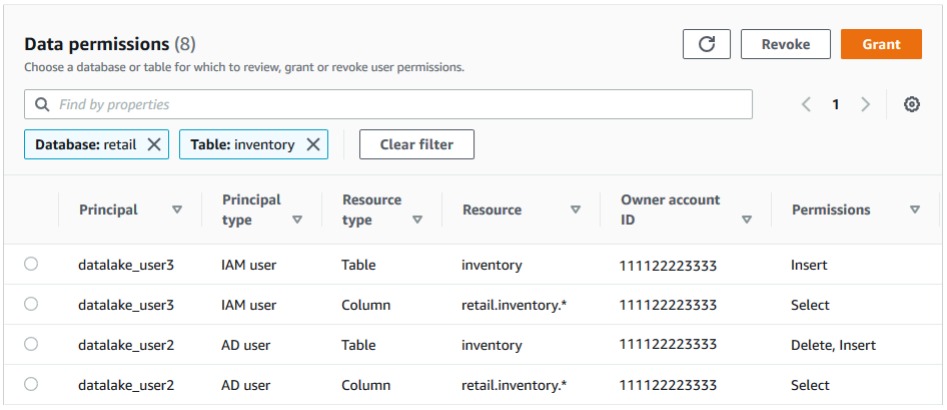
O próximo exemplo concede SELECT na tabela `inventory` usando uma lista de exclusão.

```
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"]}}}'
```

As seguintes restrições se aplicam à permissão SELECT:

- Ao conceder SELECT, você não poderá incluir a opção de concessão se a filtragem de colunas for aplicada.
- Você não pode restringir o controle de acesso em colunas que são chaves de partição.
- Uma entidade principal com a permissão SELECT em um subconjunto de colunas em uma tabela não pode receber a permissão ALTER, DROP, DELETE ou INSERT nessa tabela. Da mesma forma, uma entidade principal com a permissão ALTER, DROP, DELETE ou INSERT ou em uma tabela não pode receber a permissão SELECT com a filtragem de colunas.

A permissão SELECT sempre aparece na página Permissões de dados do console do Lake Formation como uma linha separada. A imagem a seguir mostra que SELECT é concedida aos usuários `datalake_user2` e `datalake_user3` em todas as colunas da tabela `inventory`.



Super

Permissão	Concedido neste recurso	O beneficiário também precisa
Super	DATABASE	glue:*Database*

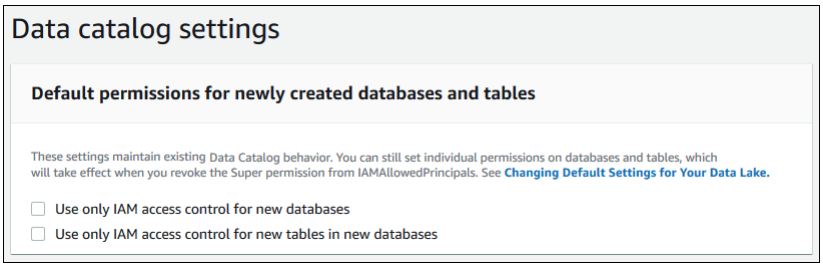
Permissão	Concedido neste recurso	O beneficiário também precisa
`Super`	`TABLE`	`glue:*Table*`, `glue:*Partition*`

Essa permissão permite que uma entidade principal execute todas as operações suportadas do Lake Formation no banco de dados ou na tabela. Você não pode conceder `Super` em um banco de dados para uma conta externa.

Essa permissão pode coexistir com as outras permissões do Lake Formation. Por exemplo, você pode conceder as permissões `Super`, `SELECT` e `INSERT` e em uma tabela de metadados. A entidade principal pode então executar todas as operações suportadas na tabela. Quando você revoga `Super`, as permissões `SELECT` e `INSERT` permanecem, e a entidade principal só pode realizar operações de seleção e inserção.

Em vez de conceder `Super` a uma entidade principal individual, você pode concedê-la ao grupo `IAMAllowedPrincipals`. O grupo `IAMAllowedPrincipals` é criado automaticamente e inclui todos os usuários e perfis do IAM que têm acesso permitido aos recursos do seu catálogo de dados por meio de suas políticas do IAM. Quando `Super` é concedido a `IAMAllowedPrincipals` para um recurso do Catálogo de sados, o acesso ao recurso é efetivamente controlado somente pelas políticas do IAM.

Você pode fazer com que a permissão `Super` seja concedida automaticamente a `IAMAllowedPrincipals` para novos recursos do catálogo aproveitando as opções na página Configurações do console do Lake Formation.



- Para conceder `Super` a `IAMAllowedPrincipals` para todos os novos bancos de dados, selecione Usar somente o controle de acesso do IAM para novos bancos de dados.
- Para conceder `Super` a `IAMAllowedPrincipals` para todas as novas tabelas em novos bancos de dados, selecione Usar somente o controle de acesso do IAM para novas tabelas em novos bancos de dados.

> **Note**
>
> Essa opção faz com que a caixa de seleção Usar somente o controle de acesso do IAM para novas tabelas nesse banco de dados na caixa de diálogo Criar banco de dados seja selecionada por padrão. Não faz nada mais do que isso. É a caixa de seleção na caixa de diálogo Criar banco de dados que permite a concessão de `Super` a `IAMAllowedPrincipals`.

Essas opções da página Configurações estão habilitadas por padrão. Para saber mais, consulte:

- the section called “Alterando as configurações padrão do seu data lake”
- the section called “Atualizar as permissões de dados AWS Glue para o modelo do Lake Formation”

SUPER_USUÁRIO

Permissão	Concedido neste recurso	O beneficiário também precisa
`Super user`	`Catalog`	`glue:GetCatalog`

É possível conceder a permissão `Super user` somente a entidades principais específicas, em catálogos dentro do Data Catalog padrão. Não é possível conceder permissão `Super user` no catálogo padrão nem em outros tipos de recurso, como bancos de dados e tabelas, ou para entidades principais em contas externas. A permissão `Super user` autoriza que uma entidade principal execute todas as operações aceitas pelo Lake Formation nos bancos de dados e nas tabelas do catálogo concedido.

Com a permissão `Super user`, a entidade principal (beneficiário) pode realizar as seguintes ações nos recursos (catálogos, bancos de dados e tabelas) do catálogo:

- Permissões `CREATE_DATABASE`, `DESCRIBE` no catálogo.
- Permissões `DROP`, `ALTER`, `CREATE_TABLE`, `DESCRIBE` (efetivamente, `SUPER`) em todos os bancos de dados do catálogo.
- Permissões `DROP`, `ALTER`, `DESCRIBE`, `SELECT`, `INSERT`, `DELETE` (efetivamente, `SUPER`) em todas as tabelas em todos os bancos de dados do catálogo.

- Permissões `All` (efetivamente, SUPER) em catálogos dentro do catálogo.
- Permissões que podem ser concedidas (a capacidade de conceder essas permissões a outras entidades principais) em todos os catálogos, bancos de dados e tabelas do catálogo.

Com a permissão `Super user` em um recurso do catálogo, o beneficiário não tem permissão para realizar nem delegar as ações `ALTER` e `DROP` no catálogo.

ASSOCIAR

Permissão	Concedido neste recurso	O beneficiário também precisa

`ASSOCIATE`	`LF-Tag`	`glue:GetDatabase`
		`glue:GetTable`
		`lakeformation:AddLFTagsToResource"`
		`lakeformation:RemoveLFTagsFromResource"`
		`lakeformation:GetResourceLFTags`
		`lakeformation:ListLFTags`
		`lakeformation:GetLFTag`
		`lakeformation:SearchTablesByLFTags`
		`lakeformation:SearchDatabasesByLFTags`

Um diretor com essa permissão em um LF-Tag pode atribuí-lo LF-Tag a um recurso do Catálogo de Dados. Conceder `ASSOCIATE` concede implicitamente `DESCRIBE`.

Example

Este exemplo concede `datalake_user1` ao usuário a `ASSOCIATE` permissão LF-Tag com a chave`module`. Ele concede permissões para visualizar e atribuir todos os valores dessa chave, conforme indicado pelo asterisco (*).

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --permissions "ASSOCIATE" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

Integrar o Centro de Identidade do IAM

Com Centro de Identidade do AWS IAM, você pode se conectar a provedores de identidade (IdPs) e gerenciar centralmente o acesso de usuários e grupos em todos os serviços de AWS análise. É possível integrar provedores de identidade, como Okta, Ping e Microsoft Entra ID (antigo Azure Active Directory), ao Centro de Identidade do IAM para que os usuários da organização acessem dados usando uma experiência de login único. O Centro de Identidade do IAM também aceita a conexão de outros provedores de identidade terceiros.

Para obter mais informações, consulte Provedores de identidade compatíveis no Guia Centro de Identidade do AWS IAM do usuário.

Você pode configurar AWS Lake Formation como um aplicativo habilitado no IAM Identity Center, e os administradores do data lake podem conceder permissões refinadas a usuários e grupos autorizados nos recursos. AWS Glue Data Catalog

Os usuários da organização podem entrar em qualquer aplicação habilitada para o Centro de Identidade usando o provedor de identidade da organização e consultar conjuntos de dados aplicando as permissões do Lake Formation. Com essa integração, você pode gerenciar o acesso aos AWS serviços sem criar várias funções do IAM.

A propagação de identidade confiável é um Centro de Identidade do AWS IAM recurso que os administradores do Connected Serviços da AWS podem usar para conceder e auditar o acesso aos dados do serviço. O acesso a esses dados é baseado em atributos do usuário, como associações de grupo. A configuração da propagação de identidade confiável requer colaboração entre os administradores do Connected Serviços da AWS e os administradores do IAM Identity Center. Para ter mais informações, consulte Prerequisites and considerations.

Para conhecer as limitações, consulte Limitações da integração com o Centro de Identidade do IAM.

Tópicos

- Pré-requisitos para integrar o Centro de Identidade do IAM ao Lake Formation
- Conectar o Lake Formation ao Centro de Identidade do IAM
- Atualizar integração com o Centro de Identidade do IAM Identity
- Excluir uma conexão do Lake Formation com o Centro de Identidade do IAM
- Conceder permissões a usuários e grupos
- Incluindo o contexto do usuário do IAM Identity Center nos CloudTrail registros

Pré-requisitos para integrar o Centro de Identidade do IAM ao Lake Formation

Veja a seguir os pré-requisitos para integrar o Centro de Identidade do IAM ao Lake Formation.

1. Habilitar o Centro de Identidade do IAM: habilitar o Centro de Identidade do IAM é um pré-requisito para oferecer compatibilidade com a autenticação e a propagação de identidade.
2. Selecionar a fonte de identidade: depois de habilitar o Centro de Identidade do IAM, é necessário ter um provedor de identificação para gerenciar usuários e grupos. É possível usar o diretório incorporado do Centro de Identidade como fonte de identidade ou usar IdP externo, como Microsoft Entra ID ou Okta.

   Para obter mais informações, consulte Gerenciar sua fonte de identidade e Conectar-se a um provedor de identidade externo no Guia Centro de Identidade do AWS IAM do usuário.
3. Crie um perfil do IAM: a função que cria a conexão do Centro de Identidade do IAM exige permissões para criar e modificar a configuração da aplicação no Lake Formation e no Centro de Identidade do IAM, conforme a política incorporada a seguir.

   É necessário adicionar permissões de acordo com as práticas recomendadas do IAM. As permissões específicas são detalhadas nos procedimentos a seguir. Para obter mais informações, consulte Getting Started with IAM Identity Center.

   JSON

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

```
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:CreateLakeFormationIdentityCenterConfiguration",
                "sso:CreateApplication",
                "sso:PutApplicationAssignmentConfiguration",
                "sso:PutApplicationAuthenticationMethod",
                "sso:PutApplicationGrant",
                "sso:PutApplicationAccessScope"
            ],
            "Resource": [
                "*"
            ]
        }
    ]
}
```

Se você estiver compartilhando recursos do Catálogo de Dados com organizações externas Contas da AWS ou organizações, deverá ter as permissões AWS Resource Access Manager (AWS RAM) para criar compartilhamentos de recursos. Para obter mais informações sobre as permissões necessárias para compartilhar recursos, consulte Pré-requisitos Cross-account de compartilhamento de dados.

As políticas incorporadas a seguir contêm permissões específicas necessárias para visualizar, atualizar e excluir propriedades da integração do Lake Formation com o Centro de Identidade do IAM.

- Use a política incorporada a seguir para que um perfil do IAM visualize uma integração do Lake Formation ao Centro de Identidade do IAM.

  JSON

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

 "lakeformation:DescribeLakeFormationIdentityCenterConfiguration",
                "sso:DescribeApplication"
```

```
            ],
            "Resource": [
                "*"
            ]
        }
    ]
}
```

- Use a política incorporada a seguir para que um perfil do IAM atualize uma integração do Lake Formation ao Centro de Identidade do IAM. A política também inclui permissões opcionais necessárias para compartilhar recursos com contas externas.

  JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:UpdateLakeFormationIdentityCenterConfiguration",

 "lakeformation:DescribeLakeFormationIdentityCenterConfiguration",
                "sso:DescribeApplication",
                "sso:UpdateApplication"
            ],
            "Resource": [
                "*"
            ]
        }
    ]
}
```

- Use a política incorporada a seguir para que um perfil do IAM exclua uma integração do Lake Formation ao Centro de Identidade do IAM.

  JSON

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

```
            "Action": [
                "lakeformation:DeleteLakeFormationIdentityCenterConfiguration",
                "sso:DeleteApplication"
            ],
            "Resource": [
                "*"
            ]
        }
    ]
}
```

- Com relação às permissões do IAM necessárias para conceder ou revogar permissões de data lake para usuários e grupos do Centro de Identidade do IAM, consulte Permissões do IAM necessárias para conceder ou revogar as permissões do Lake Formation.

Descrição das permissões

- `lakeformation:CreateLakeFormationIdentityCenterConfiguration`: cria a configuração do Lake Formation IdC.
- `lakeformation:DescribeLakeFormationIdentityCenterConfiguration`: descreve uma configuração existente do IdC.
- `lakeformation:DeleteLakeFormationIdentityCenterConfiguration`: permite excluir uma configuração existente do Lake Formation IdC.
- `lakeformation:UpdateLakeFormationIdentityCenterConfiguration`: usado para alterar uma configuração existente do Lake Formation.
- `sso:CreateApplication`: usado para criar uma aplicação IAM Identity Center.
- `sso:DeleteApplication`: usado para excluir uma aplicação IAM Identity Center.
- `sso:UpdateApplication`: usado para atualizar uma aplicação IAM Identity Center.
- `sso:PutApplicationGrant`: usado para alterar as informações do emissor de tokens confiáveis.
- `sso:PutApplicationAuthenticationMethod`: concede acesso para autenticação no Lake Formation.
- `sso:GetApplicationGrant`: usado para listar as informações do emissor de tokens confiáveis.
- `sso:DeleteApplicationGrant`: exclui as informações do emissor de tokens confiáveis.
- `sso:PutApplicationAccessScope`: adiciona ou atualiza a lista de alvos autorizados para um escopo de acesso ao Centro de Identidade do IAM para uma aplicação.

- `sso:PutApplicationAssignmentConfiguration`: usado para configurar como os usuários obtêm acesso a uma aplicação.

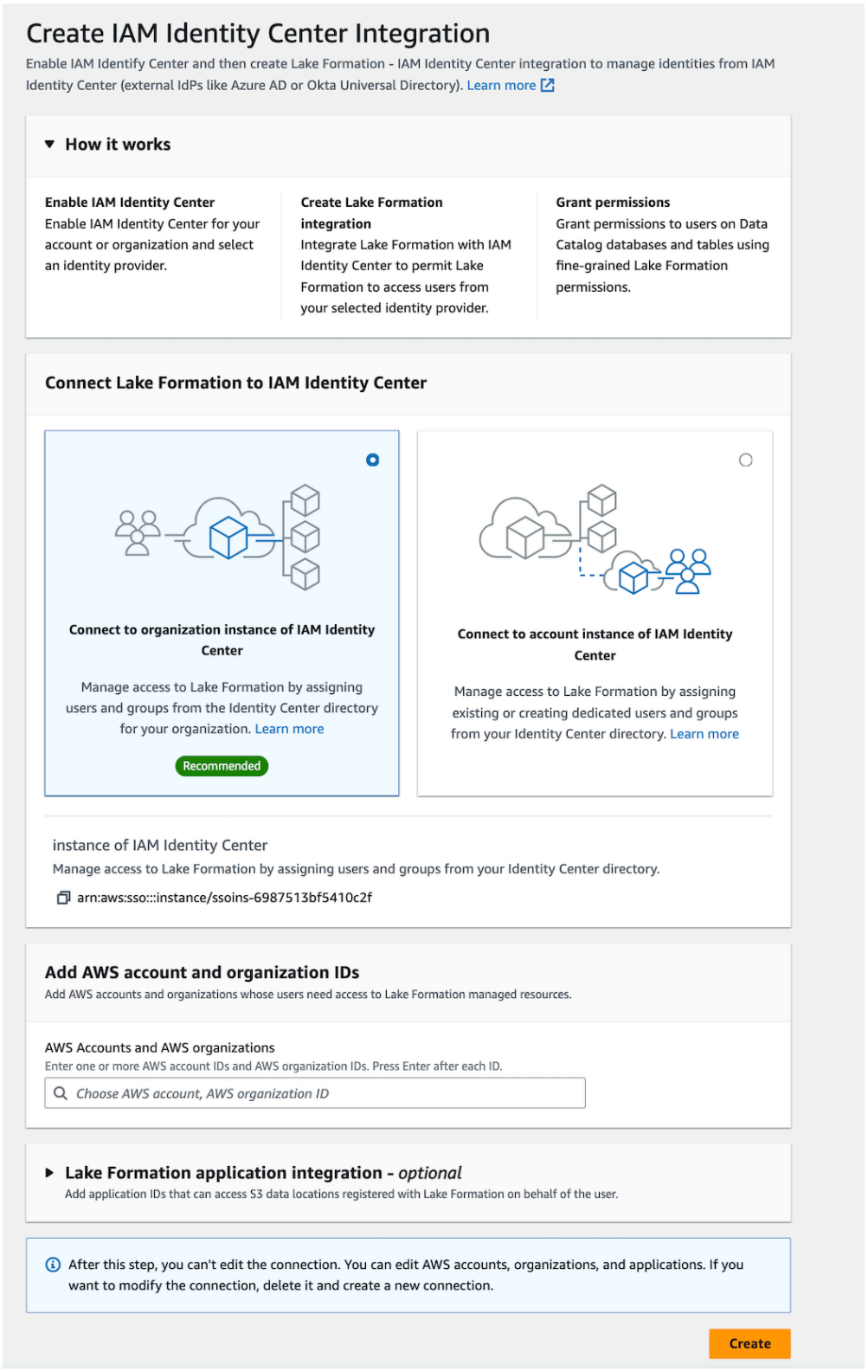
Conectar o Lake Formation ao Centro de Identidade do IAM

Antes de usar o Centro de Identidade do IAM para gerenciar identidades e conceder acesso aos recursos do catálogo de dados usando o Lake Formation, siga as etapas abaixo. É possível criar a integração do Centro de Identidade do IAM usando o console do Lake Formation ou a AWS CLI.

Console de gerenciamento da AWS

Como conectar o Lake Formation ao Centro de Identidade do IAM

1. Faça login no Console de gerenciamento da AWS, e abra o console Lake Formation em https://console.aws.amazon.com/lakeformation/.
2. No painel de navegação esquerdo, selecione Integração com o Centro de Identidade do IAM.



3. (Opcional) Insira uma ou mais Conta da AWS IDs válidas, IDs da organização e IDs da unidade and/or organizacional para permitir que contas externas acessem os recursos do Catálogo de Dados. Quando usuários ou grupos do Centro de Identidade do IAM tentam

acessar os recursos do Catálogo de Dados gerenciado do Lake Formation, o Lake Formation assume um perfil do IAM para autorizar o acesso aos metadados. Se a função do IAM pertencer a uma conta externa que não tem uma política de AWS Glue recursos e um compartilhamento de AWS RAM recursos, os usuários e grupos do IAM Identity Center não poderão acessar o recurso, mesmo que tenham permissões de Lake Formation.

O Lake Formation usa o serviço AWS Resource Access Manager (AWS RAM) para compartilhar o recurso com contas e organizações externas. AWS RAM envia um convite à conta do beneficiário para aceitar ou rejeitar o compartilhamento de recursos.

Para obter mais informações, consulte Aceitando um convite de compartilhamento de recursos do AWS RAM.

> **Note**
>
> O Lake Formation permite que os perfis do IAM de contas externas atuem como perfis operadores em nome dos usuários e grupos do Centro de Identidade do IAM para acessar os recursos do Catálogo de Dados, mas as permissões só podem ser concedidas em recursos do Catálogo de Dados dentro da conta proprietária. Se você tentar conceder permissões a usuários e grupos do IAM Identity Center em recursos do Catálogo de Dados em uma conta externa, o Lake Formation emitirá o seguinte erro: “As Cross-account concessões não são suportadas pelo diretor”.

4. (Opcional) Na tela Criar integração com o Lake Formation, especifique os ARNs de aplicações de terceiros que podem acessar dados em locais do Amazon S3 registrados no Lake Formation. A Lake Formation vende credenciais temporárias com escopo limitado na forma de AWS STS tokens para locais registrados do Amazon S3 com base nas permissões efetivas, para que aplicativos autorizados possam acessar dados em nome dos usuários.
5. (Opcional) Na tela de integração Create Lake Formation, marque a caixa de seleção Amazon Redshift Connect em Trusted Identity Propagation para permitir a descoberta de permissões federadas do Amazon Redshift via IDC. O Lake Formation propaga a identidade para sistemas downstream com base nas permissões efetivas, de modo que aplicações autorizadas possam acessar os dados em nome dos usuários.
6. Selecione Submit (Enviar).

Depois que o administrador do Lake Formation concluir as etapas e criar a integração, as propriedades do Centro de Identidade do IAM aparecerão no console do Lake Formation. A conclusão dessas tarefas torna o Lake Formation uma aplicação habilitada para o Centro de

Identidade do IAM. As propriedades no console incluem o status da integração. O status de integração indica `Success` quando está concluída. Esse status indica se a configuração do Centro de Identidade do IAM foi concluída.

AWS CLI

- O exemplo a seguir mostra como criar a integração do Lake Formation ao Centro de Identidade do IAM. Também é possível especificar o `Status` (`ENABLED`, `DISABLED`) das aplicações.

```
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"}'
```

- O exemplo a seguir mostra como visualizar a integração do Lake Formation ao Centro de Identidade do IAM.

```
aws lakeformation describe-lake-formation-identity-center-configuration
 --catalog-id <123456789012>
```

- O exemplo a seguir mostra como habilitar a `Redshift:Connect` Autorização. A autorização pode ser ATIVADA ou DESATIVADA.

```
aws lakeformation  create-lake-formation-identity-center-configuration \
--instance-arn <arn:aws:sso:::instance/ssoins-112111f12ca1122p> \
--service-integrations '[{
  "Redshift": [{
    "RedshiftConnect": {
      "Authorization": "ENABLED"
    }
  }]
}]'
```

- Use o `describe-lake-formation-identity-center-configuration` comando para descrever o aplicativo do centro de identidade de formação de lagos. `Redshift:Connecta`

integração de serviços é essencial para a propagação de identidades de IDc entre serviços e clusters:

```
aws lakeformation describe-lake-formation-identity-center-configuration --catalog-
id <123456789012>
```

Resposta:

```
{
    "CatalogId": "CATALOG ID",
    "InstanceArn": "INSTANCE ARN",
    "ApplicationArn": "APPLICATION ARN",
    "ShareRecipients": [],
    "ServiceIntegrations": [
        {
            "Redshift": [
                {
                    "RedshiftConnect": {
                        "Authorization": "ENABLED"
                    }
                }
            ]
        }
    ]
}
```

Usando o IAM Identity Center em vários Regiões da AWS

O Lake Formation oferece suporte ao IAM Identity Center em vários Regiões da AWS. Você pode estender o IAM Identity Center de sua região principal Região da AWS para outras regiões para melhorar o desempenho por meio da proximidade com os usuários e da confiabilidade. Quando uma nova região é adicionada ao IAM Identity Center, você pode criar aplicativos do Lake Formation Identity Center na nova região sem replicar identidades da região primária. Para obter mais detalhes sobre como começar a usar o IAM Identity Center em várias regiões, consulte o Multi-Region IAM Identity Center no Guia do usuário do IAM Identity Center.

Atualizar integração com o Centro de Identidade do IAM Identity

Depois de criar a conexão, é possível adicionar aplicações de terceiros para a integração com o Centro de Identidade do IAM a fim de integrá-las ao Lake Formation e obter acesso aos dados do

Amazon S3 em nome dos usuários. Também é possível remover aplicações existentes da integração com o Centro de Identidade do IAM. Você pode adicionar ou remover aplicativos usando o console Lake Formation e usando UpdateLakeFormationIdentityCenterConfiguration a operação. AWS CLI

> **Note**
>
> Depois de criar a integração com o Centro de Identidade do IAM, não é possível atualizar o ARN da instância.

Console de gerenciamento da AWS

Como atualizar uma conexão existente do Centro de Identidade do IAM com o Lake Formation

1. Faça login no Console de gerenciamento da AWS, e abra o console Lake Formation em https://console.aws.amazon.com/lakeformation/.
2. No painel de navegação esquerdo, selecione Integração com o Centro de Identidade do IAM.
3. Selecione Adicionar na página Integração com o Centro de Identidade do IAM.
4. Insira uma ou mais Conta da AWS IDs válidas, IDs da organização e IDs da unidade and/or organizacional para permitir que contas externas acessem os recursos do Catálogo de Dados.
5. Na tela Adicionar aplicações, insira os IDs das aplicações de terceiros que você deseja integrar ao Lake Formation.
6. Selecione Adicionar.
7. (Opcionalmente) Na página de integração do IAM Identity Center, você pode ativar a propagação de identidade confiável para o Amazon Redshift Connect ou desativá-la. O Lake Formation propaga a identidade para sistemas downstream com base nas permissões efetivas, de modo que aplicações autorizadas possam acessar os dados em nome dos usuários.

AWS CLI

Você pode adicionar ou remover aplicativos de terceiros para a integração do IAM Identity Center executando o AWS CLI comando a seguir. Ao definir o status de filtragem externa como ENABLED, ele permite que o Centro de Identidade do IAM forneça gerenciamento de identidade para aplicações de terceiros acessarem dados gerenciados pelo Lake Formation. Também é

possível habilitar ou desabilitar a integração com o Centro de Identidade do IAM definindo o status da aplicação.

```
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
```

Se você tem um aplicativo LF IDC existente, mas deseja adicionar a `Redshift:Connect` autorização, você pode usar o seguinte para atualizar seu aplicativo Lake Formation IDC. A autorização pode ser ATIVADA ou DESATIVADA.

```
aws lakeformation update-lake-formation-identity-center-configuration \
--service-integrations '[{

  "Redshift": [{
    "RedshiftConnect": {
      "Authorization": "ENABLED"
    }
  }]
}]'
```

Excluir uma conexão do Lake Formation com o Centro de Identidade do IAM

Se quiser excluir uma integração existente do IAM Identity Center, você pode fazer isso usando o console ou a DeleteLakeFormationIdentityCenterConfiguration operação do Lake Formation. AWS CLI

Console de gerenciamento da AWS

Como excluir uma conexão existente do Centro de Identidade do IAM com o Lake Formation

1. Faça login no Console de gerenciamento da AWS, e abra o console Lake Formation em https://console.aws.amazon.com/lakeformation/.
2. No painel de navegação esquerdo, selecione Integração com o Centro de Identidade do IAM.
3. Selecione Excluir na página Integração com o Centro de Identidade do IAM.

4. Na tela Confirmar integração, confirme a ação e selecione Excluir.

AWS CLI

Você pode excluir a integração do IAM Identity Center executando o AWS CLI comando a seguir.

```
aws lakeformation delete-lake-formation-identity-center-configuration \
    --catalog-id <123456789012>
```

Conceder permissões a usuários e grupos

O administrador do data lake pode conceder permissões a usuários e grupos do Centro de Identidade do IAM nos recursos do catálogo de dados (bancos de dados, tabelas e visualizações) para facilitar o acesso aos dados. Para conceder ou revogar as permissões do data lake, o concessor exige permissões para as ações a seguir do Centro de Identidade do IAM.

- DescribeUser
- DescribeGroup
- DescribeInstance

É possível conceder permissões usando o console do Lake Formation, a API ou a AWS CLI.

Para obter mais informações sobre a concessão de permissões, consulte the section called "Conceder permissões de dados"

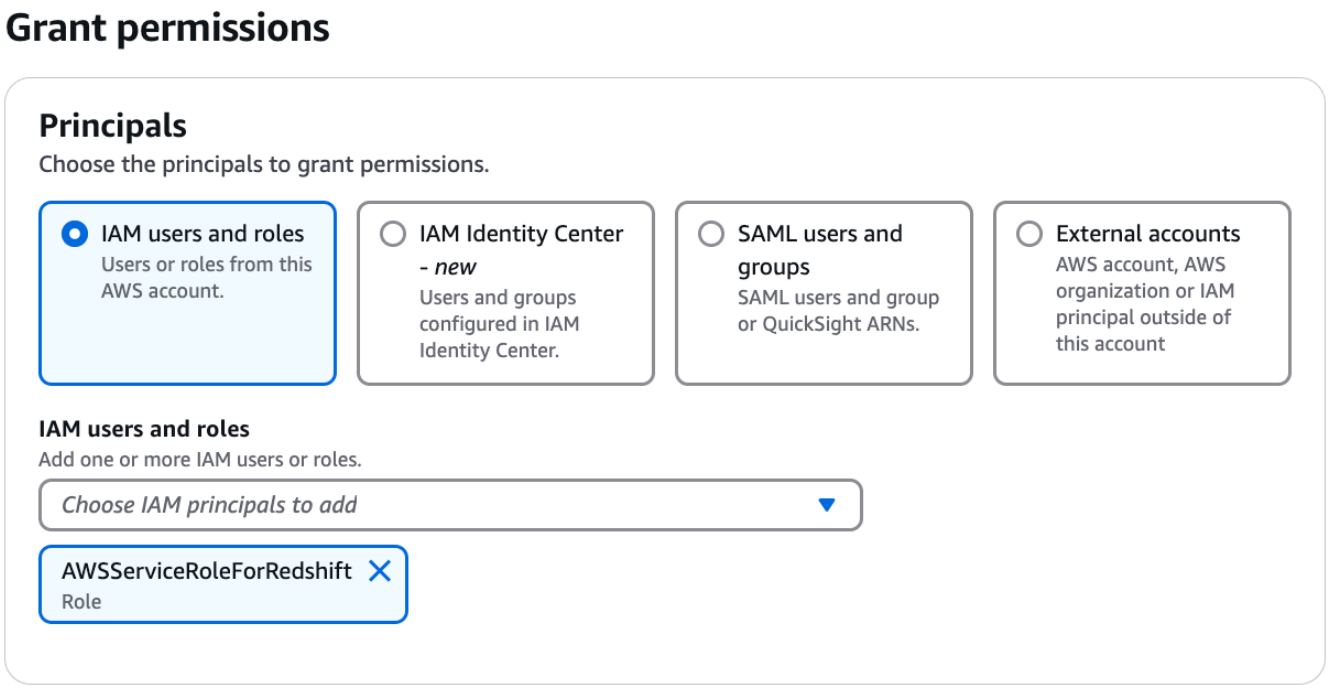
> **Note**
>
> Só é possível conceder permissões em recursos em sua conta. Para distribuir permissões em cascata para usuários e grupos em recursos compartilhados com você, você deve usar compartilhamentos de AWS RAM recursos.

Console de gerenciamento da AWS

Como conceder permissões a usuários e grupos

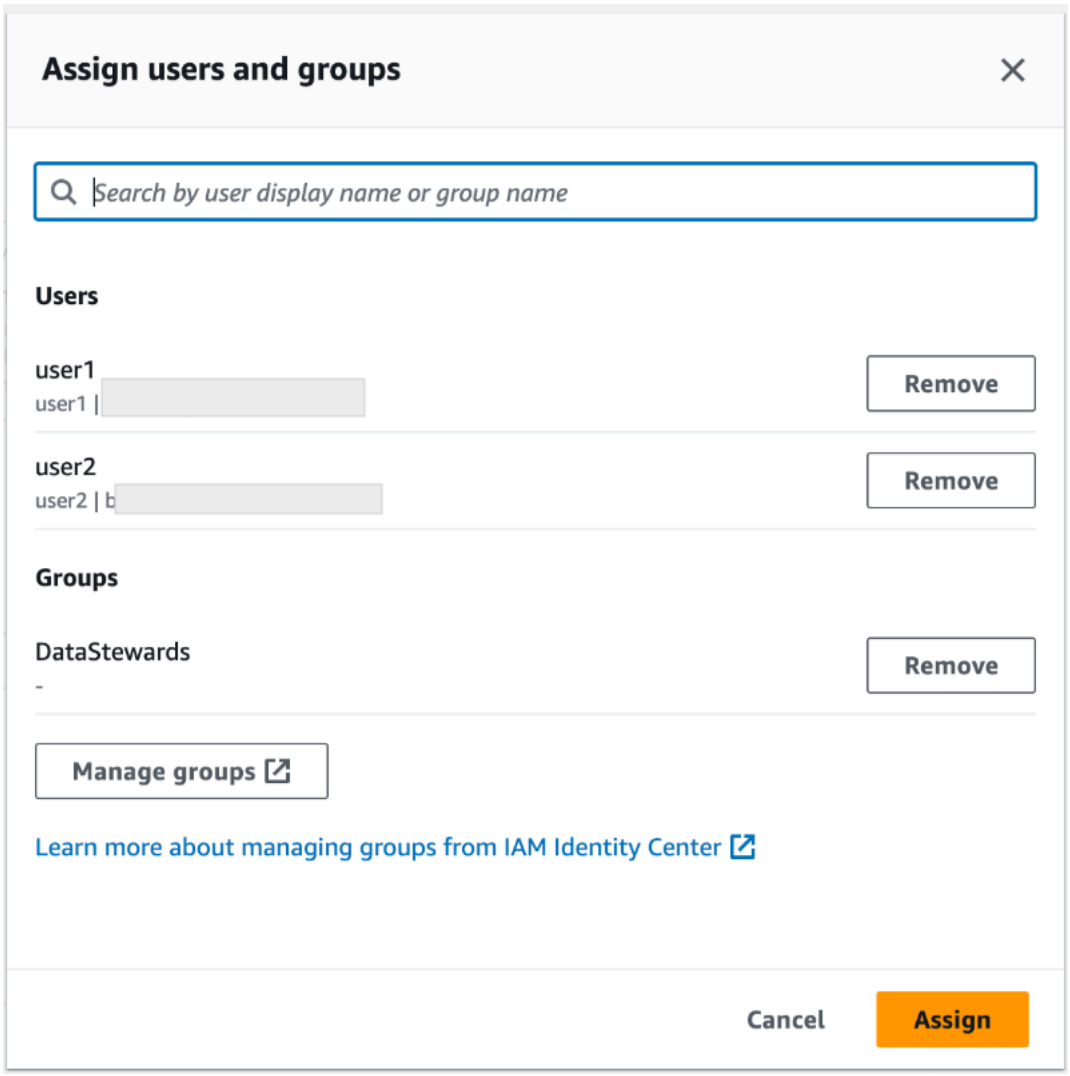
1. Faça login no Console de gerenciamento da AWS, e abra o console Lake Formation em https://console.aws.amazon.com/lakeformation/.

2. Selecione Permissões do data lake em Permissões no console do Lake Formation.
3. Selecione Conceder.
4. Na página Conceder permissões de data lake, selecione usuários e grupos do Centro de Identidade do IAM.
5. Selecione Adicionar para escolher os usuários e os grupos aos quais conceder permissões.



6. Na tela Atribuir usuários e grupos, escolha os and/or grupos de usuários para conceder permissões.

   Selecione Atribuir.



7. Depois, escolha o método para conceder permissões.

   Para obter instruções sobre como conceder permissões usando o método de recursos nomeados, consulte Conceder permissões de dados usando o método de recurso nomeado.

   Para obter instruções sobre como conceder permissão usando LF-Tags, consulteConcedendo permissões de data lake usando o método LF-TBAC.

8. Selecione os recursos do catálogo de dados nos quais deseja conceder as permissões.
9. Selecione as permissões do catálogo de dados a serem concedidas.
10. Selecione Conceder.

AWS CLI

O exemplo a seguir mostra como conceder a permissão `SELECT` ao usuário do Centro de Identidade do IAM em uma tabela.

```
aws lakeformation grant-permissions \
--principal DataLakePrincipalIdentifier=arn:aws:identitystore:::user/<UserId> \
--permissions "SELECT" \
--resource '{ "Table": { "DatabaseName": "retail", "TableWildcard": {} } }'
```

Para recuperar `UserId` do IAM Identity Center, consulte a GetUserId operação na Referência da API do IAM Identity Center.

Incluindo o contexto do usuário do IAM Identity Center nos CloudTrail registros

O Lake Formation usa a funcionalidade de fornecimento de credenciais para conceder acesso temporário aos dados do Amazon S3. Por padrão, quando um usuário do IAM Identity Center envia uma consulta a um serviço de análise integrado, os CloudTrail registros incluem apenas a função do IAM assumida pelo serviço para fornecer acesso de curto prazo. Se você usar uma função definida pelo usuário para registrar a localização de dados do Amazon S3 com o Lake Formation, poderá optar por incluir o contexto do usuário do IAM Identity Center nos CloudTrail eventos e, em seguida, rastrear os usuários que acessam seus recursos.

> **Important**
>
> Para incluir solicitações de API do Amazon S3 em nível de objeto no CloudTrail, você precisa habilitar o registro de CloudTrail eventos para buckets e objetos do Amazon S3. Para obter mais informações, consulte Habilitar o registro de CloudTrail eventos para buckets e objetos do Amazon S3 no Guia do usuário do Amazon S3.

Como habilitar a auditoria do fornecimento de credenciais em localizações de data lake registradas com perfis definidos pelo usuário

1. Faça login no console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

2. No painel de navegação à esquerda, expanda Administração e selecione Configurações do Catálogo de Dados.
3. Em Auditoria aprimorada, escolha Propagar contexto fornecido.
4. Escolha Salvar.

Você também pode ativar a opção de auditoria aprimorada definindo o `Parameters` atributo na PutDataLakeSettings operação. Por padrão, o parâmetro `SET_CONTEXT"` é definido como verdadeiro.

```
{
    "DataLakeSettings": {
        "Parameters": {"SET_CONTEXT": "true"},
    }
}
```

A seguir está um trecho de um CloudTrail evento com a opção de auditoria aprimorada. Esse log inclui o contexto de sessão do usuário do Centro de Identidade do IAM Identity e o perfil do IAM definido pelo usuário assumido pelo Lake Formation para acessar a localização de dados do Amazon S3. Veja o parâmetro `onBehalfOf` no trecho a seguir.

```
{
         "eventVersion":"1.09",
         "userIdentity":{
            "type":"AssumedRole",
            "principalId":"AROAW7F7MOX4OYE6FLIFN:access-grants-
e653760c-4e8b-44fd-94d9-309e035b75ab",
            "arn":"arn:aws:sts::123456789012:assumed-role/accessGrantsTestRole/access-
grants-e653760c-4e8b-44fd-94d9-309e035b75ab",
            "accountId":"123456789012",
            "accessKeyId":"ASIAW7F7MOX4CQLD4JIZN",
            "sessionContext":{
               "sessionIssuer":{
                  "type":"Role",
                  "principalId":"AROAW7F7MOX4OYE6FLIFN",
                  "arn":"arn:aws:iam::123456789012:role/accessGrantsTestRole",
                  "accountId":"123456789012",
                  "userName":"accessGrantsTestRole"
               },
               "attributes":{
                  "creationDate":"2023-08-09T17:24:02Z",
```

```
                "mfaAuthenticated":"false"
             }
          },
          "onBehalfOf":{
              "userId": "<identityStoreUserId>",
              "identityStoreArn": "arn:aws:identitystore::<restOfIdentityStoreArn>"
          }
       },
       "eventTime":"2023-08-09T17:25:43Z",
       "eventSource":"s3.amazonaws.com",
       "eventName":"GetObject",
....
```

Adicionar uma localização do Amazon S3 ao seu data lake

Para adicionar um local de dados como armazenamento em seu data lake, você registra o local (localização do data lake) com AWS Lake Formation. Em seguida, você pode usar as permissões do Lake Formation para um controle de acesso refinado a AWS Glue Data Catalog objetos que apontam para esse local e para os dados subjacentes no local.

O Lake Formation também permite registrar uma localização de dados no modo de acesso híbrido e fornece a flexibilidade de habilitar seletivamente as permissões do Lake Formation para bancos de dados e tabelas em seu catálogo de dados. Com o modo de acesso híbrido, você tem um caminho incremental que permite definir permissões do Lake Formation para um conjunto específico de usuários sem interromper as políticas de permissão de outros usuários ou workloads existentes.

Para obter mais informações sobre como configurar o modo de acesso híbrido, consulte Modo de acesso híbrido

Quando você registra um local, esse caminho do Amazon S3 e todas as pastas sob esse caminho são registrados.

Por exemplo, digamos que você tenha uma organização de caminhos do Amazon S3 como a seguinte:

`/mybucket/accounting/sales/`

Se você se registrar `S3://mybucket/accounting`, a pasta `sales` também será registrada e sob o gerenciamento do Lake Formation.

Para obter mais informações sobre o registro de locais, consulte Underlying data access control.

> **Note**
>
> As permissões do Lake Formation são recomendadas para dados estruturados (organizados em tabelas com linhas e colunas). Se os dados contiverem dados não estruturados baseados em objetos, pense em usar o recurso Concessão de Acesso do Amazon S3 para gerenciar o acesso aos dados.

Tópicos

- Requisitos para funções usadas para registrar locais
- Registrando uma localização do Amazon S3
- Registrando uma localização criptografada do Amazon S3
- Registrando uma localização do Amazon S3 em outra AWS account
- Registrando uma localização criptografada do Amazon S3 em AWS contas
- Cancelar o registro de uma localização do Amazon S3
- Acessando locais do Amazon S3

Requisitos para funções usadas para registrar locais

Você deve especificar uma função AWS Identity and Access Management (IAM) ao registrar uma localização do Amazon Simple Storage Service (Amazon S3). AWS Lake Formation assume essa função ao acessar os dados nesse local.

Você pode usar um dos seguintes tipos de perfil para registrar um local:

- A função vinculada ao serviço do Lake Formation. Esse perfil concede as permissões necessárias no local. O uso desse perfil é a maneira mais simples de registrar o local. Para obter mais informações, consulte Usar perfis vinculados ao serviço para o Lake Formation e Limitações de perfis vinculadas a serviços.
- Um perfil definido pelo usuário. Use um perfil definido pelo usuário quando precisar conceder mais permissões do que o perfil vinculado ao serviço fornece.

  Você deve usar um perfil definido pelo usuário nas seguintes circunstâncias:

  - Ao registrar um local em outra conta.

Para obter mais informações, consulte the section called “Registrando uma localização do Amazon S3 em outra AWS account” e the section called “Registrando uma localização criptografada do Amazon S3 em AWS contas”.

- Se você usou uma CMK AWS gerenciada (`aws/s3`) para criptografar a localização do Amazon S3.

  Para obter mais informações, consulte Registrando uma localização criptografada do Amazon S3.
- Se você planeja acessar o local usando o Amazon EMR.

  Se você já registrou um local com o perfil vinculado ao serviço e deseja começar a acessar o local com o Amazon EMR, você deve cancelar o registro do local e registrá-lo novamente com um perfil definido pelo usuário. Para obter mais informações, consulte the section called “Cancelar o registro de uma localização do Amazon S3”.

Usar perfis vinculados ao serviço para o Lake Formation

AWS Lake Formation usa uma função vinculada ao serviço AWS Identity and Access Management (IAM). Um perfil vinculado ao serviço é um tipo exclusivo de perfil do IAM vinculado diretamente ao Lake Formation. A função vinculada ao serviço é predefinida pelo Lake Formation e inclui todas as permissões que o serviço exige para chamar outros AWS serviços em seu nome.

Um perfil vinculado ao serviço facilita a configuração do Lake Formation porque você não precisa criar um perfil e adicionar manualmente as permissões necessárias. O Lake Formation define as permissões de seu perfil vinculado ao serviço e, a menos que definido de outra forma, apenas o Lake Formation pode assumir suas funções. As permissões definidas incluem a política de confiança e a política de permissões, e essa política de permissões não pode ser anexada a nenhuma outra entidade do IAM.

Este perfil vinculado ao serviço confia nos seguintes serviços para assumir a função:

- `lakeformation.amazonaws.com`

Quando você usa um perfil vinculado ao serviço na conta A para registrar uma localização do Amazon S3 que é de propriedade da conta B, a política de bucket do Amazon S3 (uma política baseada em recursos) na conta B deve conceder permissões de acesso ao perfil vinculado ao serviço na conta A.

Para acessar informações sobre o uso de perfis vinculados a serviços para registrar um local de dados, consulte Limitações de perfis vinculadas a serviços.

> **Note**
>
> As políticas de controle de serviço (SCPs) não afetam as funções vinculadas ao serviço. Para obter mais informações, consulte Políticas de controle de serviço (SCPs) no guia AWS Organizations do usuário.

Permissões de perfil vinculado ao serviço para o Lake Formation

O Lake Formation usa o perfil vinculado ao serviço chamado `AWSServiceRoleForLakeFormationDataAccess`. Essa função fornece um conjunto de permissões do Amazon Simple Storage Service (Amazon S3) que permitem que o serviço integrado Lake Formation (Amazon Athena como) acesse locais registrados. Ao registrar um local de data lake, você deve fornecer uma função que tenha as read/write permissões necessárias do Amazon S3 nesse local. Em vez de criar um perfil com as permissões necessárias para o Amazon S3, você pode usar esse perfil vinculado ao serviço.

Na primeira vez que você nomeia o perfil vinculado ao serviço como o perfil com o qual registrar um caminho, o perfil vinculado ao serviço e uma nova política do IAM são criadas em seu nome. O Lake Formation adiciona o caminho à política embutida e o anexa ao perfil vinculado ao serviço. Quando você registra caminhos subsequentes com o perfil vinculado ao serviço, o Lake Formation adiciona o caminho à política existente.

Enquanto estiver conectado como administrador do data lake, registre um local do data lake. Em seguida, no console do IAM, pesquise o perfil `AWSServiceRoleForLakeFormationDataAccess` e visualize as políticas anexadas.

Por exemplo, depois de registrar o local `s3://my-kinesis-test/logs`, o Lake Formation cria a seguinte política embutida e a anexa a `AWSServiceRoleForLakeFormationDataAccess`.

JSON

```
{
    "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"
            ]
        }
    ]
}
```

Criar um perfil vinculado a serviços para o Lake Formation

Não é necessário criar manualmente um perfil vinculado ao serviço. Quando você registra um local do Amazon S3 com o Lake Formation na Console de gerenciamento da AWS, na ou na AWS API AWS CLI, o Lake Formation cria a função vinculada ao serviço para você.

> **Important**
>
> Esse perfil vinculado ao serviço pode aparecer em sua conta se você concluiu uma ação em outro serviço que usa os atributos compatíveis com esse perfil. Para saber mais, consulte Uma nova função apareceu na minha conta do IAM.

Se excluir essa função vinculada ao serviço e precisar criá-la novamente, você pode usar esse mesmo processo para recriar a função na sua conta. Quando você registra uma localização do Amazon S3 com o Lake Formation, o Lake Formation cria o perfil vinculado ao serviço para você outra vez.

Também é possível usar o console do IAM para criar um perfil vinculado ao serviço com o caso de uso do Lake Formation. Na AWS CLI ou na AWS API, crie uma função vinculada ao serviço com o nome do `lakeformation.amazonaws.com` serviço. Para saber mais, consulte Criar um perfil vinculado a serviço no Guia do usuário do IAM. Se você excluir essa função vinculada ao serviço, será possível usar esse mesmo processo para criar a função novamente.

Editar um perfil vinculado a serviços para o Lake Formation

O Lake Formation não permite que você edite o perfil vinculado ao serviço `AWSServiceRoleForLakeFormationDataAccess`. Depois que criar um perfil vinculado ao serviço, você não poderá alterar o nome do perfil, pois várias entidades podem fazer referência a ele. No entanto, será possível editar a descrição do perfil usando o IAM. Para saber mais, consulte Editar uma função vinculada a serviço no Guia do usuário do IAM.

Excluir um perfil vinculado a serviços para o Lake Formation

Se você não precisar mais usar um recurso ou serviço que requer um perfil vinculado ao serviço, é recomendável excluí-lo. Dessa forma, você não tem uma entidade não utilizada que não seja monitorada ativamente ou mantida. No entanto, você deve limpar os recursos de seu perfil vinculado ao serviço antes de excluí-lo manualmente.

> **Note**
>
> Se o serviço Lake Formation estiver usando o perfil quando você tentar excluir os recursos, a exclusão poderá falhar. Se isso acontecer, espere alguns minutos e tente a operação novamente.

Para excluir os recursos do Lake Formation usados pelo Lake Formation

- Se você usou o perfil vinculado ao serviço para registrar localizações do Amazon S3 com o Lake Formation, antes de excluí-lo, será necessário cancelar o registro da localização e registrá-la novamente usando um perfil personalizado.

Como excluir manualmente o perfil vinculado ao serviço usando o IAM

Use o console do IAM AWS CLI, o ou a AWS API para excluir a função `AWSServiceRoleForLakeFormationDataAccess` vinculada ao serviço. Para saber mais, consulte Excluir um perfil vinculado ao serviço no Guia do usuário do IAM.

A seguir estão os requisitos para um perfil definido pelo usuário:

- Ao criar o novo perfil, na página Criar perfil do console do IAM, escolha Serviço da AWS e, em seguida, em Escolha um caso de uso, escolha Lake Formation.

  Se você criar o perfil usando um caminho diferente, certifique-se de que o perfil tenha uma relação de confiança com `lakeformation.amazonaws.com`. Para acessar mais informações, consulte Modificar uma política de confiança de perfil (console).
- A função deve ter uma política em linha que conceda ao Amazon read/write S3 permissões no local. A seguir está uma política típica.

  JSON

```
{
    "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"
            ]
        }
```

```
    ]
}
```

- Adicione a política de confiança a seguir ao perfil do IAM para permitir que o serviço Lake Formation assuma o perfil e forneça credenciais temporárias aos mecanismos de analytics integrados.

  Para incluir o contexto do usuário do IAM Identity Center nos CloudTrail registros, a política de confiança deve ter a permissão para a `sts:SetContext` ação.

  JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "DataCatalogViewDefinerAssumeRole1",
            "Effect": "Allow",
            "Principal": {
               "Service": [
                    "lakeformation.amazonaws.com"
                ]
            },
            "Action": [
                "sts:AssumeRole",
                "sts:SetContext"
            ]
        }
    ]
}
```

- O administrador do data lake que registra o local deve ter a permissão `iam:PassRole` sobre o perfil.

  A seguir está uma política embutida que concede essa permissão. *<account-id>*Substitua por um número de AWS conta válido e *<role-name>* substitua pelo nome da função.

  JSON

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

```
            "Sid": "PassRolePermissions",
            "Effect": "Allow",
            "Action": [
                "iam:PassRole"
            ],
            "Resource": [
                "arn:aws:iam::111122223333:role/<role-name>"
            ]
        }
    ]
}
```

- Para permitir que o Lake Formation adicione CloudWatch registros em Logs e publique métricas, adicione a seguinte política em linha.

> **Note**
>
> A gravação no CloudWatch Logs incorre em uma cobrança.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "Sid1",
            "Effect": "Allow",
            "Action": [
                "logs:CreateLogStream",
                "logs:CreateLogGroup",
                "logs:PutLogEvents"
            ],
            "Resource": [
                 "arn:aws:logs:us-east-1:111122223333:log-group:/aws-
lakeformation-acceleration/*",
                 "arn:aws:logs:us-east-1:111122223333:log-group:/aws-
lakeformation-acceleration/*:log-stream:*"
            ]
        }
    ]
```

```
}
```

Registrando uma localização do Amazon S3

Você deve especificar uma função AWS Identity and Access Management (IAM) ao registrar uma localização do Amazon Simple Storage Service (Amazon S3). O Lake Formation assume essa função quando concede credenciais temporárias a AWS serviços integrados que acessam os dados naquele local.

> **Important**
>
> Evite registrar um bucket do Amazon S3 que tenha o Solicitante paga ativado. Para buckets registrados no Lake Formation, a função usada para registrar o bucket é sempre vista como solicitante. Se o bucket for acessado por outra AWS conta, o proprietário do bucket será cobrado pelo acesso aos dados se a função pertencer à mesma conta do proprietário do bucket.

Você pode usar o AWS Lake Formation console, a API Lake Formation ou AWS Command Line Interface (AWS CLI) para registrar uma localização no Amazon S3.

Antes de começar

Analise os requisitos da função usada para registrar o local.

Para registrar uma localização (console)

> **Important**
>
> Os procedimentos a seguir pressupõem que a localização do Amazon S3 esteja na mesma AWS conta do Catálogo de Dados e que os dados na localização não estejam criptografados. Outras seções deste capítulo abrangem o registro de várias contas e o registro de locais criptografados.

1. Abra o AWS Lake Formation console em https://console.aws.amazon.com/lakeformation/. Faça login como administrador do data lake ou como usuário com a permissão `lakeformation:RegisterResource` do IAM.

2. No painel de navegação, em Administração selecione Localizações do data lake.
3. Escolha Registrar localização e, em seguida, escolha Procurar para selecionar um caminho do Amazon Simple Storage Service (Amazon S3).
4. (Opcional, mas altamente recomendado) Selecione Revisar permissões de local para ver uma lista de todos os recursos existentes no local selecionado do Amazon S3 e as permissões.

   Registrar o local selecionado pode fazer com que seus usuários do Lake Formation tenham acesso aos dados que já estão nesse local. A visualização dessa lista ajuda a garantir que os dados existentes permaneçam seguros.
5. Para o perfil do IAM, escolha a função `AWSServiceRoleForLakeFormationDataAccess` vinculada ao serviço (a padrão) ou um perfil personalizado do IAM que atenda aos requisitos em the section called “Requisitos para funções usadas para registrar locais”.

   É possível atualizar um local registrado ou outros detalhes somente ao registrá-lo usando um perfil personalizado do IAM. Para editar um local registrado usando um perfil vinculado ao serviço, é necessário cancelar o registro do local e registrá-lo novamente.
6. Escolha a opção Ativar Federação do Catálogo de Dados para permitir que o Lake Formation assuma uma função e forneça credenciais temporárias aos AWS serviços integrados para acessar tabelas em bancos de dados federados. Se um local estiver registrado no Lake Formation e você quiser usar o mesmo local para uma tabela em um banco de dados federado, será necessário registrar o mesmo local com a opção Habilitar federação do catálogo de dados.
7. Escolha o Modo de acesso híbrido para não ativar as permissões do Lake Formation por padrão. Ao registrar o local do Amazon S3 no modo de acesso híbrido, você pode habilitar as permissões do Lake Formation optando por entidades principais para bancos de dados e tabelas nesse local.

   Para obter mais informações sobre como configurar o modo de acesso híbrido, consulte Modo de acesso híbrido.
8. Escolha Registrar caminho do S3 com a conta do proprietário e forneça a conta do proprietário do bucket do S3 para permitir que o Lake Formation verifique o acesso da função de registro no local do Amazon S3 e permita que o Lake Formation venda credenciais temporárias para arquivos do Amazon S3 no local registrado.
9. Selecione Registrar local.

Para registrar um local (AWS CLI)

1. Registrar o novo local no Lake Formation

   Este exemplo usa um perfil vinculado ao serviço para registrar o local. Em vez disso, você pode usar o argumento `--role-arn` para fornecer sua própria função.

   *<s3-path>*Substitua por um caminho válido do Amazon S3, o número da conta por uma AWS conta válida e *<s3-access-role>* por uma função do IAM que tenha permissões para registrar um local de dados.

   > **Note**
   >
   > Não será possível editar propriedades de um local registrado se ele estiver registrado usando um perfil vinculado ao serviço.

   ```
   aws lakeformation register-resource \
    --resource-arn arn:aws:s3:::<s3-path> \
    --use-service-linked-role
   ```

   O exemplo a seguir usa um perfil personalizado para registrar o local.

   ```
   aws lakeformation register-resource \
    --resource-arn arn:aws:s3:::<s3-path> \
    --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role>
   ```

2. Como atualizar o local registrado no Lake Formation

   Será possível editar um local registrado somente se ele estiver registrado usando um perfil personalizado do IAM. Para um local registrado com um perfil vinculado ao serviço, é necessário cancelar o registro do local e registrá-lo novamente. Para obter mais informações, consulte the section called “Cancelar o registro de uma localização do 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. Registrar um local de dados no modo de acesso híbrido com federação

```
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
```

4. Registre um local com a conta de proprietário esperada

   Use a `--expected-resource-owner-account` bandeira para permitir que o Lake Formation verifique o acesso da função de registro no local do Amazon S3 e venda credenciais temporárias para arquivos do Amazon S3 no local registrado.

```
aws lakeformation register-resource \
 --resource-arn arn:aws:s3:::<s3-path> \
 --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role> \
 --expected-resource-owner-account <123456789012>
```

```
aws lakeformation update-resource \
 --resource-arn arn:aws:s3:::<s3-path> \
 --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role> \
 --expected-resource-owner-account <123456789012>
```

Para obter mais informações, consulte Operação RegisterResourceda API.

> **Note**
>
> Depois de registrar uma localização no Amazon S3, qualquer AWS Glue tabela apontando para a localização (ou qualquer uma de suas localizações secundárias) retornará o valor do `IsRegisteredWithLakeFormation` parâmetro como `true` na `GetTable` chamada. Há uma limitação conhecida de que as operações da API do catálogo de dados, como `GetTables` e `SearchTables`, não atualizam o valor do parâmetro `IsRegisteredWithLakeFormation` e retornam o padrão, que é falso. É recomendável usar a API `GetTable` para visualizar o valor correto do parâmetro `IsRegisteredWithLakeFormation`.

Registrando uma localização criptografada do Amazon S3

O Lake Formation se integra com AWS Key Management Service (AWS KMS) para permitir que você configure com mais facilidade outros serviços integrados para criptografar e descriptografar dados em locais do Amazon Simple Storage Service (Amazon S3).

Tanto o cliente é gerenciado AWS KMS keys Chaves gerenciadas pela AWS quanto o suporte. Atualmente, o lado do cliente encryption/decryption é compatível somente com o Athena.

Você deve especificar uma função AWS Identity and Access Management (IAM) ao registrar uma localização no Amazon S3. Para locais criptografados do Amazon S3, a função deve ter permissão para criptografar e descriptografar dados com o. Ou a política de chaves do AWS KMS key KMS deve conceder permissões sobre a chave da função.

> **Important**
>
> Evite registrar um bucket do Amazon S3 que tenha o Solicitante paga ativado. Para buckets registrados no Lake Formation, a função usada para registrar o bucket é sempre vista como solicitante. Se o bucket for acessado por outra AWS conta, o proprietário do bucket será cobrado pelo acesso aos dados se a função pertencer à mesma conta do proprietário do bucket.

O Lake Formation usa um perfil vinculado a serviços para registrar seus locais de dados. No entanto, esse perfil tem várias limitações. Devido a essas restrições, recomendamos criar e usar um perfil

personalizado do IAM para ter maior flexibilidade e controle. O perfil personalizado que você cria para registrar o local deve atender aos requisitos especificados em the section called “Requisitos para funções usadas para registrar locais”.

> **Important**
>
> Se você usou um Chave gerenciada pela AWS para criptografar a localização do Amazon S3, você não pode usar a função vinculada ao serviço Lake Formation. Você deve usar um papel personalizado e adicionar permissões do IAM na chave do perfil. Os detalhes são fornecidos posteriormente nesta seção.

Os procedimentos a seguir explicam como registrar um local do Amazon S3 criptografado com uma chave gerenciada pelo cliente ou uma Chave gerenciada pela AWS.

- Registrar um local criptografado com uma chave gerenciada pelo cliente
- Registrando um local criptografado com um Chave gerenciada pela AWS

Antes de começar

Analise os requisitos da função usada para registrar o local.

Para registrar uma localização do Amazon S3 criptografada com uma chave gerenciada pelo cliente

> **Note**
>
> Se a chave KMS ou a localização do Amazon S3 não estiverem na AWS mesma conta do catálogo de dados, siga as instruções the section called “Registrando uma localização criptografada do Amazon S3 em AWS contas” em vez disso.

1. Abra o AWS KMS console em https://console.aws.amazon.com/kmse faça login como um usuário administrativo AWS Identity and Access Management (IAM) ou como um usuário que pode modificar a política de chaves da chave KMS usada para criptografar o local.
2. No painel de navegação, selecione Chaves gerenciadas pelo cliente e selecione o nome da chave do KMS desejada.

3. Na página de detalhes da chave KMS, escolha a guia Política de chaves e, em seguida, faça o seguinte para adicionar sua função personalizada ou a função vinculada ao serviço Lake Formation como usuário da chave KMS:

   - Se a visualização padrão estiver sendo exibida (com as seções Administradores de chaves, Exclusão de chaves, Usuários de chaves e Outras AWS contas), na seção Usuários principais, adicione sua função personalizada ou a função vinculada ao serviço Lake Formation. `AWSServiceRoleForLakeFormationDataAccess`
   - Se a política de chaves (JSON) estiver sendo exibida – edite a política para adicionar sua função personalizada ou a função `AWSServiceRoleForLakeFormationDataAccess` vinculada ao serviço Lake Formation ao objeto “Permitir o uso da chave”, conforme mostrado no exemplo a seguir.

   > **Note**
   >
   > Se esse objeto estiver ausente, adicione-o com as permissões mostradas no exemplo. O exemplo usa a função vinculada ao serviço.

```
        ...
        {
            "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. Abra o AWS Lake Formation console em https://console.aws.amazon.com/lakeformation/. Faça login como administrador do data lake ou como usuário com a permissão `lakeformation:RegisterResource` do IAM.
5. No painel de navegação, em Administração em Locais de data lake.
6. Escolha Registrar localização e, em seguida, escolha Procurar para selecionar um caminho do Amazon Simple Storage Service (Amazon S3).
7. (Opcional, mas altamente recomendado) Escolha Revisar permissões de localização para ver uma lista de todos os recursos existentes no local selecionado do Amazon S3 e suas permissões.

   Registrar o local selecionado pode fazer com que seus usuários do Lake Formation tenham acesso aos dados que já estão nesse local. A visualização dessa lista ajuda a garantir que os dados existentes permaneçam seguros.
8. Para o perfil do IAM, escolha a função `AWSServiceRoleForLakeFormationDataAccess` vinculada ao serviço (a padrão) ou sua função personalizada que atenda a the section called "Requisitos para funções usadas para registrar locais".
9. Escolha Registrar local.

Para saber mais sobre a função vinculada ao serviço do , consulte Permissões de perfil vinculado ao serviço para o Lake Formation.

Para registrar uma localização do Amazon S3 criptografada com um Chave gerenciada pela AWS

> **Important**
>
> Se a localização do Amazon S3 não estiver na mesma AWS conta do catálogo de dados, siga as instruções em the section called "Registrando uma localização criptografada do Amazon S3 em AWS contas" vez disso.

1. Crie um perfil do IAM para usar para registrar o local. Certifique-se de que ele atenda aos requisitos listados em the section called "Requisitos para funções usadas para registrar locais".
2. Adicione a seguinte política em linha à função. Ele concede permissões sobre a chave da função. A especificação `Resource` deve designar o nome do recurso da Amazon (ARN) da

Chave gerenciada pela AWS. Você pode obter o ARN no AWS KMS console. Para obter o ARN correto, certifique-se de fazer login no AWS KMS console com a mesma AWS conta e região Chave gerenciada pela AWS que foram usadas para criptografar o local.

JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "kms:Encrypt",
        "kms:Decrypt",
        "kms:ReEncrypt*",
        "kms:GenerateDataKey*",
        "kms:DescribeKey"
      ],
      "Resource": "arn:aws:kms:us-
east-1:111122223333:key/1234abcd-12ab-34cd-56ef-1234567890ab"
    }
  ]
}
```

Você pode usar aliases de chave do KMS em vez do ID da chave: `arn:aws:kms:region:account-id:key/alias/your-key-alias`.

Para obter mais informações, consulte Aliases na AWS KMS seção do Guia do AWS Key Management Service desenvolvedor.

3. Abra o AWS Lake Formation console em https://console.aws.amazon.com/lakeformation/. Faça login como administrador do data lake ou como usuário com a permissão `lakeformation:RegisterResource` do IAM.
4. No painel de navegação, em Administração em Locais de data lake.
5. Escolha Registrar localização e, em seguida, escolha Procurar para selecionar um caminho do Amazon S3.
6. (Opcional, mas altamente recomendado) Escolha Revisar permissões de localização para ver uma lista de todos os recursos existentes no local selecionado do Amazon S3 e suas permissões.

Registrar o local selecionado pode fazer com que seus usuários do Lake Formation tenham acesso aos dados que já estão nesse local. A visualização dessa lista ajuda a garantir que os dados existentes permaneçam seguros.

7. Em Perfil do IAM, escolha a função que você criou na Etapa 1.
8. Escolha Registrar local.

Registrando uma localização do Amazon S3 em outra AWS account

AWS Lake Formation permite que você registre localizações do Amazon Simple Storage Service (Amazon S3) em todas as contas. AWS Por exemplo, se AWS Glue Data Catalog estiver na conta A, um usuário na conta A poderá registrar um bucket do Amazon S3 na conta B.

Registrar um bucket do Amazon S3 AWS na conta B usando AWS Identity and Access Management uma função (IAM) AWS na conta A requer as seguintes permissões:

- O papel na conta A deve conceder permissões no bucket na conta B.
- A política de bucket na conta B deve conceder permissões de acesso à função na conta A.

> **Important**
>
> Evite registrar um bucket do Amazon S3 que tenha o Solicitante paga ativado. Para buckets registrados no Lake Formation, a função usada para registrar o bucket é sempre vista como solicitante. Se o bucket for acessado por outra AWS conta, o proprietário do bucket será cobrado pelo acesso aos dados se a função pertencer à mesma conta do proprietário do bucket.
>
> Você não pode usar a função vinculada ao serviço Lake Formation para registrar um local em outra conta. Em vez disso, é necessário usar uma função definida pelo usuário. A função deve atender aos requisitos do the section called “Requisitos para funções usadas para registrar locais”. Para saber mais sobre a função vinculada ao serviço do , consulte Permissões de perfil vinculado ao serviço para o Lake Formation.

Antes de começar

Analise os requisitos da função usada para registrar o local.

Para registrar um local em outro AWS account

> **Note**
>
> Se o local estiver criptografado, siga as instruções em the section called “Registrando uma localização criptografada do Amazon S3 em AWS contas” vez disso.

O procedimento a seguir pressupõe que uma entidade principal na conta 1111-2222-3333, que contém o catálogo de dados, queira registrar o bucket `awsexamplebucket1` do Amazon S3, que está na conta 1234-5678-9012.

1. Na conta 1111-2222-3333, faça login Console de gerenciamento da AWS e abra o console do IAM em. https://console.aws.amazon.com/iam/
2. Crie uma nova função ou visualize uma função existente que atenda aos requisitos de the section called “Requisitos para funções usadas para registrar locais”. Certifique-se de que a função concede permissões do Amazon S3 em `awsexamplebucket1`.
3. Abra o console do Amazon S3 em https://console.aws.amazon.com/s3/. Faça login com a conta 1234-5678-9012.
4. Na lista Nome do bucket, escolha o nome do bucket, `awsexamplebucket1`.
5. Escolha Permissões.
6. Na página Permissões, escolha Política de bucket.
7. No Editor de políticas do bucket, cole a política a seguir. Substitua *<role-name>* pelo nome da sua função.

   JSON

```
{
    "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. Escolha Salvar.
9. Abra o AWS Lake Formation console em https://console.aws.amazon.com/lakeformation/. Faça login na conta 1111-2222-3333 como administrador do data lake ou como usuário com permissões suficientes para registrar locais.
10. No painel de navegação, em Administração em Locais de data lake.
11. Na página Locais de data lake, selecione Registrar local.
12. Na página Registrar localização, para o caminho do Amazon S3, insira o nome `s3://awsexamplebucket1` do bucket.

> **Note**
>
> Você deve digitar o nome do bucket porque os buckets de várias contas não aparecem na lista quando você escolhe Procurar.

13. Para o perfil do IAM, escolha seu perfil.
14. Escolha Registrar local.

Registrando uma localização criptografada do Amazon S3 em AWS contas

AWS Lake Formation se integra com AWS Key Management Service(AWS KMS) para permitir que você configure com mais facilidade outros serviços integrados para criptografar e descriptografar dados em locais do Amazon Simple Storage Service (Amazon S3).

Tanto as chaves gerenciadas pelo cliente quanto Chaves gerenciadas pela AWS as são suportadas. Client-side encryption/decryption não é suportado.

> **Important**
>
> Evite registrar um bucket do Amazon S3 que tenha o Solicitante paga ativado. Para buckets registrados no Lake Formation, a função usada para registrar o bucket é sempre vista como solicitante. Se o bucket for acessado por outra AWS conta, o proprietário do bucket será cobrado pelo acesso aos dados se a função pertencer à mesma conta do proprietário do bucket.

Esta seção explica como registrar uma localização do Amazon S3 nas seguintes circunstâncias:

- Os dados no local do Amazon S3 são criptografados com uma chave KMS criada no AWS KMS.
- A localização do Amazon S3 não está na mesma AWS conta do. AWS Glue Data Catalog
- A chave KMS está ou não na mesma AWS conta do Catálogo de Dados.

O registro de um bucket do Amazon S3 AWS KMS criptografado AWS na conta B usando AWS Identity and Access Management uma função (IAM) AWS na conta A requer as seguintes permissões:

- O papel na conta A deve conceder permissões no bucket na conta B.
- A política de bucket na conta B deve conceder permissões de acesso à função na conta A.
- Se a chave KMS estiver na conta B, a política de chaves deverá conceder acesso à função na conta A, e a função na conta A deverá conceder permissões na chave KMS.

No procedimento a seguir, você cria uma função na AWS conta que contém o Catálogo de Dados (conta A na discussão anterior). Em seguida, você usa essa função para registrar o local. O Lake Formation assume essa função ao acessar dados subjacentes no Amazon S3. A função assumida tem as permissões necessárias na chave do KMS. Como resultado, você não precisa conceder permissões na chave KMS às entidades principais que acessam dados subjacentes com trabalhos de ETL ou com serviços integrados, como o Amazon Athena.

> **Important**
>
> Você não pode usar a função vinculada ao serviço Lake Formation para registrar um local em outra conta. Em vez disso, é necessário usar uma função definida pelo usuário. A função deve atender aos requisitos do the section called “Requisitos para funções usadas para registrar locais”. Para saber mais sobre a função vinculada ao serviço do , consulte Permissões de perfil vinculado ao serviço para o Lake Formation.

Antes de começar

Analise os requisitos da função usada para registrar o local.

Para registrar uma localização criptografada do Amazon S3 em AWS contas

1. Na mesma AWS conta do Catálogo de dados, faça login Console de gerenciamento da AWS e abra o console do IAM emhttps://console.aws.amazon.com/iam/.
2. Crie uma nova função ou visualize uma função existente que atenda aos requisitos de the section called “Requisitos para funções usadas para registrar locais”. Certifique-se de que a função inclua uma política que concede permissões do Amazon S3 no local.
3. Se a chave KMS não estiver na mesma conta do Catálogo deDados, adicione à função uma política em linha que conceda as permissões necessárias na chave KMS. Veja abaixo um exemplo de política . Substitua Região e ID da conta pela região e o número da conta da chave do KMS. *<key-id>*Substitua pelo ID da chave.

   JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
        "Effect": "Allow",
        "Action": [
            "kms:Encrypt",
            "kms:Decrypt",
            "kms:ReEncrypt*",
            "kms:GenerateDataKey*",
            "kms:DescribeKey"
         ],
```

```
        "Resource": "arn:aws:kms:us-east-1:111122223333:key/<key-id>"
        }
    ]
}
```

4. No console do Amazon S3, adicione uma política de bucket concedendo as permissões do Amazon S3 necessárias para a função. A seguir há um exemplo de política de bucket. Substitua o ID da AWS conta pelo número da conta do Catálogo de Dados, pelo nome da sua função e *<bucket-name>* pelo nome do bucket. *<role-name>*

   JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect":"Allow",
            "Principal": {
                "AWS":"arn:aws:iam::111122223333:role/<role-name>"
            },
            "Action":"s3:ListBucket",
            "Resource":"arn:aws:s3:::<bucket-name>"
        },
        {
            "Effect":"Allow",
            "Principal": {
                "AWS":"arn:aws:iam::111122223333:role/<role-name>"
            },
            "Action": [
                "s3:DeleteObject",
                "s3:GetObject",
                "s3:PutObject"
            ],
            "Resource":"arn:aws:s3:::<bucket-name>/*"
        }
    ]
}
```

5. Em AWS KMS, adicione a função como usuário da chave KMS.

a. Abra o AWS KMS console em https://console.aws.amazon.com/kms. Em seguida, faça login como usuário administrador ou como usuário que pode modificar a política de chaves da chave KMS usada para criptografar o local.

b. No painel de navegação, selecione Chaves gerenciadas pelo cliente e selecione o nome da chave do KMS.

c. Na página de detalhes da chave KMS, na guia Política de chaves, se a visualização JSON da política de chaves não estiver sendo exibida, escolha Alternar para visualização de política.

d. Na seção Política de chaves, escolha Editar e adicione o nome do recurso da Amazon (ARN) da função ao objeto `Allow use of the key`, conforme mostrado no exemplo a seguir.

> **Note**
>
> Se esse objeto estiver ausente, adicione-o com as permissões mostradas no exemplo.

```
...
{
    "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": "*"
},
...
```

Para obter mais informações, consulte Permitir que usuários de outras contas usem uma chave KMS no Guia do desenvolvedor do AWS Key Management Service .

6. Abra o AWS Lake Formation console em https://console.aws.amazon.com/lakeformation/. Faça login na conta AWS do catálogo de dados como administrador do data lake.
7. No painel de navegação, em Administração em Locais de data lake.
8. Escolha Registrar local.
9. Na página Registrar localização, para caminho do Amazon S3, insira o caminho da localização como **`s3://<bucket>/<prefix>`**. *`<bucket>`*Substitua pelo nome do bucket e *`<prefix>`* pelo resto do caminho do local.

> **Note**
>
> Você deve digitar o caminho porque os buckets entre contas não aparecem na lista quando você escolhe Procurar.

10. Para o perfil do IAM, escolha a função na Etapa 2.
11. Escolha Registrar local.

Cancelar o registro de uma localização do Amazon S3

Você pode cancelar o registro de uma localização do Amazon Simple Storage Service (Amazon S3) se não quiser mais que ela seja gerenciada pelo Lake Formation. O cancelamento do registro de um local não afeta as permissões de localização de dados do Lake Formation concedidas nesse local. Você pode registrar novamente um local que você cancelou e as permissões de localização de dados permanecerão em vigor. Você pode usar uma função diferente para registrar novamente o local.

Para cancelar o registro de um local (console)

1. Abra o AWS Lake Formation console em https://console.aws.amazon.com/lakeformation/. Faça login como administrador do data lake ou como usuário com a permissão `lakeformation:RegisterResource` do IAM.
2. No painel de navegação, em Administração em Locais de data lake.
3. Selecione um local e, no menu Ações, escolha Remover.

4. Quando solicitada a confirmação, escolha Remover.

Acessando locais do Amazon S3

AWS Lake Formation permite que você leia e grave os arquivos de dados subjacentes no Amazon Simple Storage Service (Amazon S3) para tabelas registradas AWS Glue Data Catalog no (Catálogo de dados). Isso fornece um único conjunto de permissões para consultas SQL e acesso direto a arquivos usando suas concessões de tabela existentes do Lake Formation. A Lake Formation estende sua funcionalidade de venda de credenciais para locais do Amazon S3 registrados no catálogo de dados. Quando seu aplicativo solicita acesso aos arquivos do Amazon S3 subjacentes a uma tabela do Catálogo de Dados, o Lake Formation avalia as permissões existentes em nível de tabela do chamador e, se autorizado, retorna credenciais do Amazon S3 com escopo curto para a localização registrada dessa tabela.

Habilite o acesso direto aos locais do Amazon S3

Para ativar esse recurso, você deve concluir as etapas a seguir.

Pré-requisitos

Preencha os seguintes pré-requisitos antes de ativar o sinalizador de recurso:

- Especifique a localização da tabela — Especifique o caminho para seus dados no Amazon S3 usando a propriedade LOCATION nos atributos da tabela. Para obter mais informações sobre o formato de localização, consulte Localização da tabela no Amazon S3 no Guia do usuário do Amazon Athena.
- Registre locais do Amazon S3 — Catalogue suas tabelas no catálogo de dados de seus locais do Amazon S3 e registre-as no Lake Formation fornecendo a conta do proprietário do bucket do Amazon S3. Quando você registra um local, esse caminho do Amazon S3 e todas as pastas sob esse caminho são registrados. Para obter mais informações, consulte Registrando uma localização do Amazon S3.
- Conceda permissões do Lake Formation — Conceda permissões do Lake Formation `SELECT` (ou`SUPER`) em tabelas para seus cientistas de dados e aplicativos usando o console ou as APIs do Lake Formation. Se você já usa o Lake Formation para acessar tabelas com serviços como Athena ou Amazon EMR, isso já está configurado.
- Habilitar integração de aplicativos com acesso total à tabela — Ative a integração de aplicativos com acesso total à tabela para permitir a venda de credenciais para o local da tabela registrada. Para obter mais informações, consulte Integração de aplicativos para acesso total à tabela.

Ativar o sinalizador de recurso

Ative a `fs.s3a.lakeformation.access.grants.enabled` bandeira nos mecanismos compatíveis:

```
fs.s3a.lakeformation.access.grants.enabled = true
```

- Amazon EMR no Amazon EC2 — EMR 7.13 ou posterior
- Amazon EMR no EKS — EMR 7.13 ou posterior
- Amazon EMR Serverless — EMR 7.13 ou posterior

Depois de configurado, você pode ler ou gravar imediatamente arquivos de dados do EMR usando APIs padrão.

Para obter mais informações sobre a integração com o Amazon EMR, consulte a venda de credenciais baseada em caminhos do Lake Formation no Guia de gerenciamento do Amazon EMR.

Como funciona

O acesso à localização do Lake Formation-based Amazon S3 segue esse fluxo:

1. Um diretor ou função solicita acesso aos arquivos de dados do Amazon S3 por meio de um serviço, como um trabalho do Amazon EMR Spark ou um pipeline de processamento de dados.
2. O plug-in AWS SDK integrado ao EMR intercepta a solicitação do Amazon S3 e chama a API Lake Formation. GetTemporaryDataLocationCredentials
3. O Lake Formation verifica se o caminho do Amazon S3 corresponde a um local registrado e a uma tabela catalogada e avalia as permissões do chamador na tabela associada. AWS Glue
4. Se o chamador tiver `SELECT` ou tiver `SUPER` permissões na tabela correspondente com acesso total à tabela, o Lake Formation vende credenciais temporárias com escopo específico para esse local do Amazon S3. `SELECT`concede credenciais READ; `SUPER` concede credenciais READWRITE.
5. O plug-in usa essas credenciais para concluir a solicitação do Amazon S3, e o Amazon S3 fornece os dados para o aplicativo.
6. Todas as operações de venda automática de credenciais são registradas CloudTrail, fornecendo uma trilha de auditoria para acesso aos dados.

Esse processo é transparente para você e você simplesmente usará as permissões existentes do Lake Formation com APIs padrão baseadas em arquivos.

```
# Read raw data (Lake Formation-based S3 Location access)
raw_df = spark.read.json("s3://finance-datalake/raw/transactions/dt=2024-03-21/")

# Read governed data (Lake Formation-based S3 Location access)
transactions_df = spark.read.parquet("s3://data-lake/transactions/year=2026/")

# Write processed data (Lake Formation-based S3 Location access)
processed_df.write \
    .mode("append") \
    .partitionBy("transaction_date") \
    .parquet("s3://finance-datalake/processed/transactions/")

print("ETL complete. Records written:", processed_df.count())
```

Requisitos de permissão

Atualmente, o recurso vende credenciais somente quando o chamador tem acesso total à tabela, ou seja, SELECT em todas as colunas sem filtros de linha ou coluna. Se uma tabela tiver filtros em nível de linha ou coluna aplicados, os chamadores devem continuar usando mecanismos confiáveis, como Athena, Amazon EMR AWS Glue ou Amazon Redshift, que podem aplicar esses filtros. Isso garante que os limites de segurança permaneçam consistentes.

Cross-account acesso

O acesso simplificado do Lake Formation aos locais do Amazon S3 funciona com recursos de compartilhamento entre contas. Quando você compartilha uma tabela com outra AWS conta por meio do Lake Formation, os destinatários podem acessar os arquivos de dados subjacentes do Amazon S3 usando suas permissões do Lake Formation, sujeitos aos mesmos requisitos completos de acesso à tabela.

O recurso oferece suporte ao mecanismo de link de recursos do Lake Formation para acesso entre contas. Quando uma conta de consumidor tem permissões em uma tabela compartilhada, o Lake Formation vende credenciais com escopo definido para a localização do Amazon S3 da conta registrada, permitindo acesso contínuo aos dados entre contas sem exigir políticas separadas de bucket do Amazon S3 ou funções do IAM entre contas.

Locais aninhados do Amazon S3

Quando várias tabelas apontam para locais aninhados no mesmo bucket, o Lake Formation aplica o seguinte comportamento:

- Ao acessar`s3://bucket`, você recebe permissões correspondentes à tabela registrada no nível do bucket.
- Ao acessar`s3://bucket/folder1`, você recebe permissões correspondentes à tabela registrada nesse caminho específico.
- Para acessar pastas sem uma tabela registrada (por exemplo,`s3://bucket/folder2`), você recebe permissões do local registrado principal mais próximo.
- Se várias tabelas forem registradas no mesmo local, o Lake Formation retornará um erro devido a permissões conflitantes.

Auditoria e conformidade

Todas as operações de venda automática de credenciais são registradas CloudTrail, fornecendo uma trilha de auditoria para acesso aos dados. Quando o Lake Formation vende credenciais por meio da `GetTemporaryDataLocationCredentials` API, CloudTrail registra:

- O principal (usuário ou função)
- Timestamp
- Local do Amazon S3
- AWS GlueTabela associada

As chamadas subsequentes da API do Amazon S3 feitas com essas credenciais também são registradas como eventos de dados do Amazon CloudTrail S3, com o contexto vinculando-as à concessão de permissão do Lake Formation. Isso fornece aos auditores uma visibilidade completa sobre quem acessou quais dados, quando e por qual caminho de permissão, tudo em um único CloudTrail fluxo de log.

Compatível AWS serviços

O acesso simplificado do Lake Formation aos locais do Amazon S3 funciona com:

- Amazon EMR no Amazon EC2 (EMR 7.13 ou posterior)
- Amazon EMR no EKS (EMR 7.13 ou posterior)
- Amazon EMR Serverless (EMR 7.13 ou posterior)

Plugin de código aberto para serviços de terceiros

Você também pode integrar seus Trino aplicativos Apache Spark ou aplicativos usando APIs ou por meio de um plug-in de código aberto fornecido pela AWS. Para obter mais informações, consulte aws-lakeformation-accessgrants-plugin-java-v2 em. GitHub

Considerações

Observe as seguintes considerações ao usar a venda de credenciais para locais do Amazon S3:

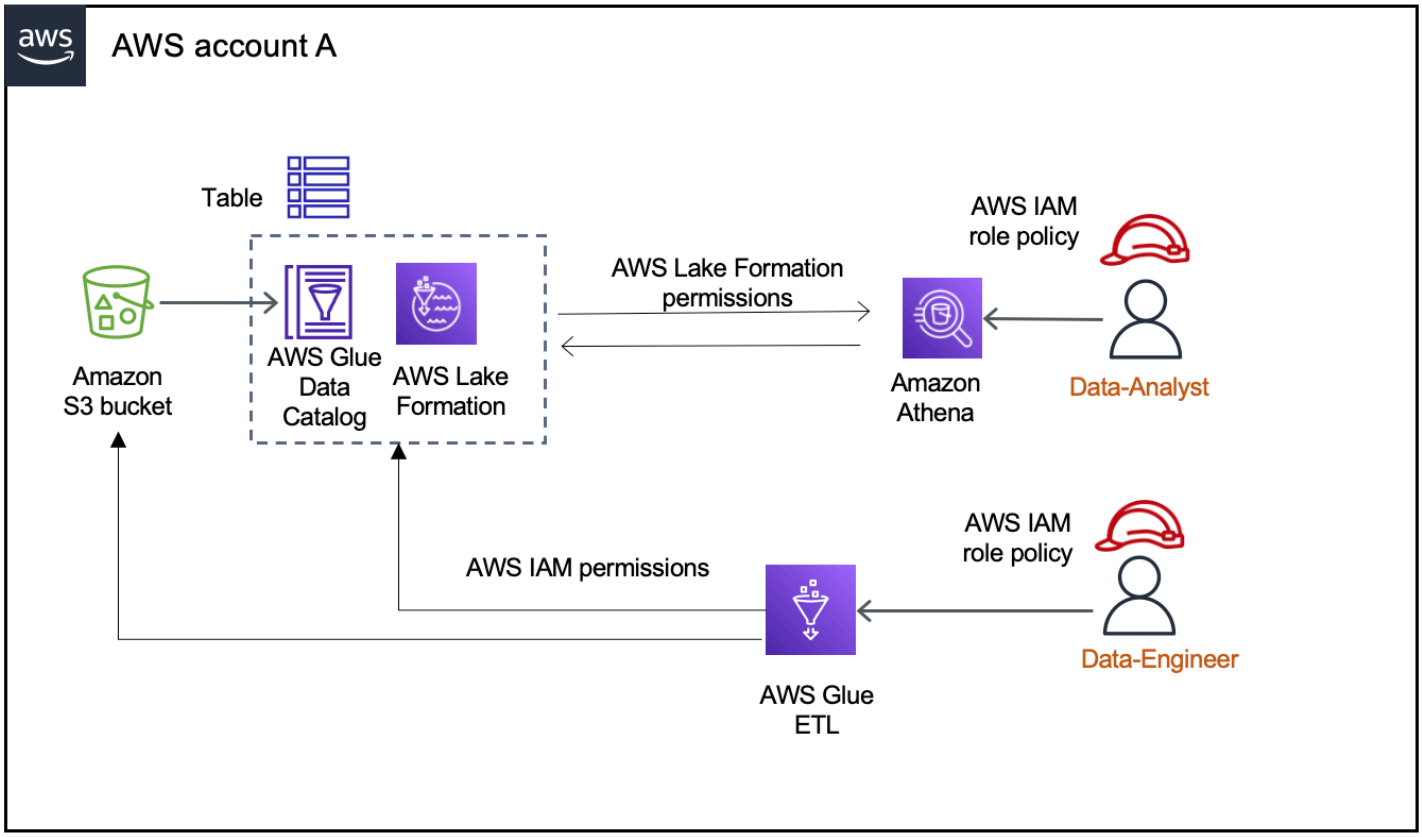
- A venda de credenciais para locais do Amazon S3 não é suportada entre regiões.
- A venda de credenciais é suportada para locais do Amazon S3 incluídos como o local principal da tabela.
- O plug-in suporta Apache Hive formatos de Delta Lake tabela e. Apache Hudi Apache Icebergnão é suportado atualmente.
- Atualmente, o plug-in não é compatível com o modo Amazon EMR Spark Fine-Grained Access Control (FGAC).

Modo de acesso híbrido

AWS Lake Formation o modo de acesso híbrido suporta dois caminhos de permissão para os mesmos AWS Glue Data Catalog objetos.
No primeiro caminho, o Lake Formation permite que você selecione entidades principais específicas e conceda a elas permissões do Lake Formation para acessar catálogos, bancos de dados, tabelas e visualizações, se você optar por isso. O segundo caminho permite que todos os outros diretores acessem esses recursos por meio das políticas principais padrão do IAM para Amazon AWS Glue S3 e ações.

Ao registrar um local do Amazon S3 com o Lake Formation, você tem a opção de aplicar permissões do Lake Formation para todos os recursos desse local ou usar o modo de acesso híbrido. O modo de acesso híbrido impõe somente `CREATE_TABLE`, `CREATE_PARTITION`, `UPDATE_TABLE` permissões por padrão. Quando um local do Amazon S3 está no modo híbrido, você pode habilitar as permissões do Lake Formation optando pelas entidades principais para os objetos do Data Catalog nesse local. Isso significa que tanto as permissões do Lake Formation quanto as permissões do IAM podem controlar o acesso a esses dados. Isso significa que as entidades principais aceitas precisarão tanto das permissões do Lake Formation quanto das permissões do IAM para acessar os dados, enquanto as entidades principais não aceitas continuarão acessando os dados apenas com as permissões do IAM.

Assim, o modo de acesso híbrido oferece a flexibilidade de habilitar seletivamente o Lake Formation para catálogos, bancos de dados e tabelas em seu Data Catalog para um conjunto específico de usuários sem interromper o acesso de outros usuários ou workloads existentes.



Para ver as considerações e as limitações, consulte Considerações e limitações do modo de acesso híbrido.

Termos e definições

Aqui estão as definições dos recursos do catálogo de dados com base em como você configura as permissões de acesso:

Recurso do Lake Formation

Um recurso registrado no Lake Formation. Os usuários precisam de permissões do Lake Formation para acessar o recurso.

AWS Glue recurso

Um recurso registrado no Lake Formation. Os usuários precisam apenas de permissões do IAM para acessar o recurso porque ele tem permissões de grupo `IAMAllowedPrincipals`. As permissões do Lake Formation não são aplicadas.

Para obter mais informações sobre as permissões de grupo `IAMAllowedPrincipals`, consulte Permissões de metadados.

Recurso híbrido

Um recurso registrado no modo de acesso híbrido. Com base nos usuários que acessam o recurso, ele alterna dinamicamente entre ser um recurso do Lake Formation ou um recurso do AWS Glue .

Casos de uso comuns do modo de acesso híbrido

Você pode usar o modo de acesso híbrido para fornecer acesso em cenários de compartilhamento de dados com uma única conta e entre contas:

Cenários de conta única

- Converter um AWS Glue recurso em um recurso híbrido — Nesse cenário, você não está usando o Lake Formation no momento, mas deseja adotar as permissões do Lake Formation para objetos do Catálogo de Dados. Ao registrar o local do Amazon S3 no modo de acesso híbrido, você pode conceder permissões do Lake Formation aos usuários que optam por bancos de dados e tabelas específicos que apontam para esse local.
- Converter um recurso do Lake Formation em um recurso híbrido: no momento, você está usando as permissões do Lake Formation para controlar o acesso a um banco de dados do Catálogo de Dados, mas deseja fornecer acesso a novas entidades principais usando permissões do IAM para o Amazon S3 e o AWS Glue sem interromper as permissões existentes do Lake Formation.

  Quando você atualiza um registro de localização de dados para o modo de acesso híbrido, novas entidades principais podem acessar o banco de dados do catálogo de dados apontando a localização do Amazon S3 usando políticas de permissões do IAM sem interromper as permissões Lake Formation dos usuários existentes.

  Antes de atualizar o registro de localização de dados para ativar o modo de acesso híbrido, você precisa primeiro optar pelas entidades principais que atualmente estão acessando o recurso com as permissões do Lake Formation.
  Isso evita possíveis interrupções no fluxo de trabalho atual.
  Você também precisa conceder permissão `Super` nas tabelas do banco de dados ao grupo `IAMAllowedPrincipal`.

Cross-account cenários de compartilhamento de dados

- Compartilhe AWS Glue recursos usando o modo de acesso híbrido — Nesse cenário, a conta do produtor tem tabelas em um banco de dados que atualmente são compartilhadas com uma conta de consumidor usando políticas de permissões do IAM para Amazon S3 e AWS Glue ações. A localização dos dados do banco de dados não está registrada no Lake Formation.

  Antes de registrar a localização dos dados no modo de acesso híbrido, você precisa atualizar as Configurações da versão entre contas para a versão 4. A versão 4 fornece as novas políticas de AWS RAM permissão necessárias para o compartilhamento entre contas quando o `IAMAllowedPrincipal` grupo tem `Super` permissão sobre o recurso. Para esses recursos com permissões de grupo `IAMAllowedPrincipal`, você pode conceder permissões do Lake Formation para contas externas e permitir que eles usem as permissões do Lake Formation. O administrador do data lake na conta do destinatário pode conceder permissões do Lake Formation às entidades principais da conta e autorizá-las a aplicar as permissões do Lake Formation.
- Compartilhe recursos do Lake Formation usando o modo de acesso híbrido – Atualmente, a conta de produtor tem tabelas em um banco de dados que são compartilhadas com uma conta de consumidor que impõe as permissões do Lake Formation. A localização dos dados do banco de dados é registrada no Lake Formation.

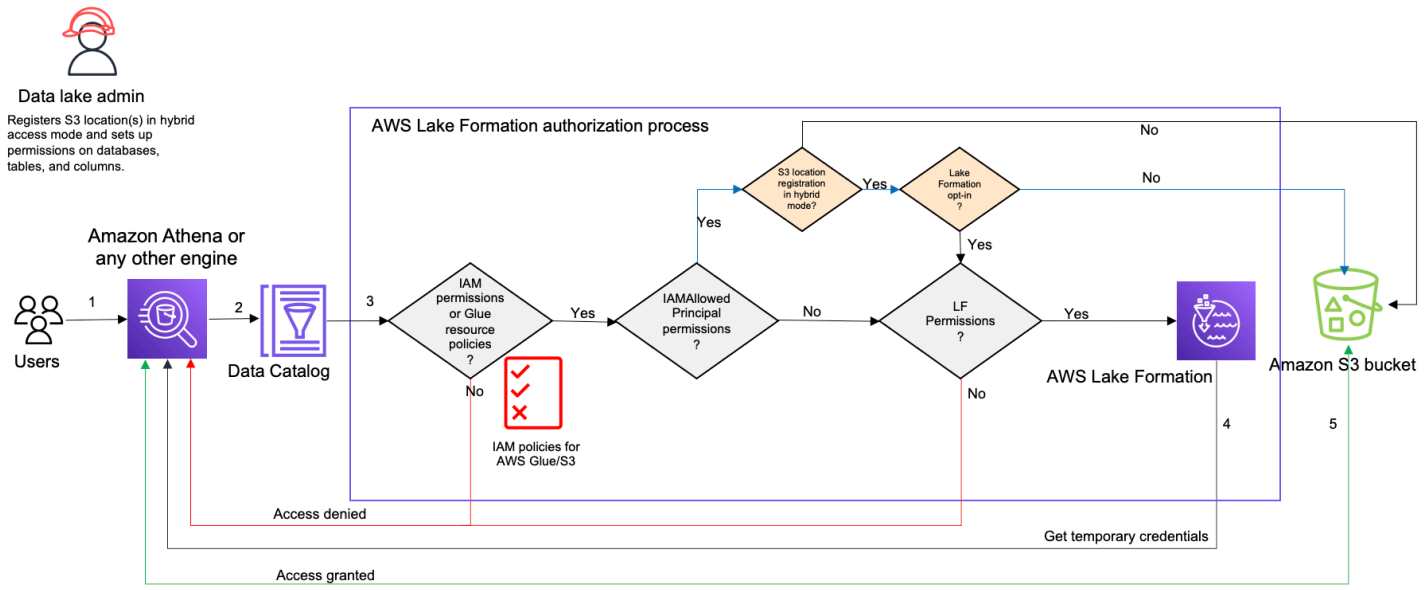
  Nesse caso, você pode atualizar o registro de localização do Amazon S3 para o modo de acesso híbrido e compartilhar os dados do Amazon S3 e os metadados do catálogo de dados usando as políticas de bucket do Amazon S3 e as políticas de recursos do catálogo de dados com as entidades principais na conta do consumidor. Você precisa conceder novamente as permissões existentes do Lake Formation e optar pelas entidades principais antes de atualizar o registro de localização do Amazon S3. Além disso, você precisa conceder permissão `Super` nas tabelas do banco de dados ao grupo `IAMAllowedPrincipals`.

Tópicos

- Como funciona o modo de acesso híbrido
- Configurando o modo de acesso híbrido - cenários comuns
- Removendo entidades principais e recursos do modo de acesso híbrido
- Visualizando entidades principais e recursos no modo de acesso híbrido
- Recursos adicionais do

Como funciona o modo de acesso híbrido

O diagrama a seguir mostra como a autorização do Lake Formation funciona no modo de acesso híbrido quando você consulta os recursos do catálogo de dados.



Antes de acessar os dados em seu data lake, um administrador de data lake ou um usuário com permissões administrativas configura políticas de usuário individuais da tabela do catálogo de dados para permitir ou negar o acesso às tabelas em seu catálogo de dados. Em seguida, uma entidade principal que tem as permissões para realizar a operação `RegisterResource` registra a localização da tabela no Amazon S3 com o Lake Formation no modo de acesso híbrido. Se um local de dados não estiver registrado no Lake Formation, o administrador vai conceder permissões do Lake Formation a usuários específicos nos bancos de dados e tabelas do Data Catalog e os autorizar a usar as permissões do Lake Formation para esses bancos de dados e tabelas no modo de acesso híbrido.

1. Envia uma consulta - Um principal envia uma consulta ou um script de ETL usando um serviço integrado, como Amazon Athena, Amazon EMR ou Amazon Redshift AWS Glue Spectrum.
2. Solicita dados - O mecanismo analítico integrado identifica a tabela que está sendo solicitada e envia a solicitação de metadados para o catálogo de dados (`GetTable`, `GetDatabase`).
3. Verifica as permissões - O catálogo de dados verifica as permissões de acesso da entidade principal consultora com o Lake Formation.
   a. Se a tabela não tiver permissões de grupo `IAMAllowedPrincipals` anexadas, as permissões do Lake Formation serão aplicadas.

b. Se a entidade principal tiver optado por usar as permissões do Lake Formation no modo de acesso híbrido e a tabela tiver permissões do grupo `IAMAllowedPrincipals` anexadas, as permissões do Lake Formation serão aplicadas. O mecanismo de consulta aplica os filtros recebidos do Lake Formation e retorna os dados ao usuário.

c. Se a localização da tabela não estiver registrada no Lake Formation e a entidade principal não tiver optado por usar as permissões do Lake Formation no modo de acesso híbrido, o catálogo de dados verificará se a tabela tem permissões do grupo `IAMAllowedPrincipals` anexadas a ela. Se essa permissão existir na tabela, todas as entidades principais da conta receberão permissões `Super` ou `All` na tabela.

   A venda de credenciais do Lake Formation não está disponível, mesmo quando aceita, a menos que o local dos dados esteja registrado no Lake Formation.

4. Obter credenciais – O catálogo de dados verifica e informa ao mecanismo se a localização da tabela está registrada no Lake Formation ou não. Se os dados subjacentes estiverem registrados no Lake Formation, o mecanismo analítico solicitará ao Lake Formation credenciais temporárias para acessar os dados no bucket do Amazon S3.
5. Obter dados – Se a entidade principal estiver autorizada a acessar os dados da tabela, o Lake Formation fornecerá acesso temporário ao mecanismo analítico integrado. Ao usar o acesso temporário, o mecanismo analítico busca os dados do Amazon S3 e executa a filtragem necessária, como filtragem por coluna, linha ou célula. Quando o mecanismo termina de executar o trabalho, ele retorna os resultados para o usuário. Esse processo chamado de fornecimento de credenciais. Para obter mais informações, consulte Integração com o Lake Formation.
6. Se a localização dos dados da tabela não estiver registrada no Lake Formation, a segunda chamada do mecanismo analítico será feita diretamente para o Amazon S3. A política de bucket do Amazon S3 e a política de usuário do IAM em questão são avaliadas para acesso aos dados. Sempre que você usar as políticas do IAM, siga as práticas recomendadas do IAM. Para obter mais informações, consulte Práticas recomendadas de segurança no IAM no Guia do usuário do IAM.

Configurando o modo de acesso híbrido - cenários comuns

Assim como nas permissões do Lake Formation, você geralmente tem dois tipos de cenários nos quais pode usar o modo de acesso híbrido para gerenciar o acesso aos dados: fornecer acesso aos principais dentro de um Conta da AWS e fornecer acesso a um externo Conta da AWS ou principal.

Esta seção fornece instruções para configurar o modo de acesso híbrido nos seguintes cenários:

Gerencie permissões no modo de acesso híbrido em um Conta da AWS

- Convertendo um AWS Glue recurso para um recurso híbrido— No momento, você está fornecendo acesso às tabelas em um banco de dados para todos os diretores da sua conta usando as permissões do IAM para o Amazon S3 AWS Glue , mas deseja adotar o Lake Formation para gerenciar as permissões de forma incremental.

- Convertendo um recurso do Lake Formation em um recurso híbrido: no momento, você está usando o Lake Formation para gerenciar o acesso às tabelas em um banco de dados para todas as entidades principais da sua conta, mas deseja usar o Lake Formation somente para entidades principais específicas. Você deseja fornecer acesso a novos diretores usando as permissões do IAM para o Amazon S3 no mesmo banco de dados AWS Glue e tabelas.

Gerencie permissões no modo de acesso híbrido em Conta da AWS s

- Compartilhando um AWS Glue recurso usando o modo de acesso híbrido: no momento, você não está usando o Lake Formation para gerenciar as permissões de uma tabela, mas deseja aplicar as permissões do Lake Formation para fornecer acesso às entidades principais em outra conta.
- Compartilhando um recurso do Lake Formation usando o modo de acesso híbrido— Você está usando o Lake Formation para gerenciar o acesso a uma tabela, mas deseja fornecer acesso para diretores em outra conta usando permissões do IAM para o Amazon S3 no mesmo banco de dados AWS Glue e tabelas.

Configurando o modo de acesso híbrido — High-level etapas

1. Registre a localização dos dados do Amazon S3 com o Lake Formation selecionando o Modo de acesso híbrido.
2. As entidades principais devem ter permissão `DATA_LOCATION` em um local de data lake para criar tabelas ou bancos de dados do catálogo de dados que apontem para esse local.
3. Defina a configuração da Cross-account versão para a Versão 4.
4. Conceda permissões refinadas a usuários ou perfis específicos do IAM em bancos de dados e tabelas. Ao mesmo tempo, certifique-se de definir permissões `Super` ou `All` para o grupo `IAMAllowedPrincipals` no banco de dados e para todas ou para as tabelas selecionadas no banco de dados.

5. Opte pelas entidades principais e recursos. Outros diretores da conta podem continuar acessando os bancos de dados e tabelas usando as políticas de permissão do IAM AWS Glue e as ações do Amazon S3.
6. Opcionalmente, limpe as políticas de permissão do IAM para o Amazon S3 para as entidades principais que optaram por usar as permissões do Lake Formation.

Pré-requisitos para configurar o modo de acesso híbrido

Estes estão os pré-requisitos para configurar o modo de acesso híbrido:

> **Note**
>
> Recomendamos que um administrador do Lake Formation registre a localização do Amazon S3 no modo de acesso híbrido e opte por entidades principais e recursos.

1. Conceda permissão de localização de dados (`DATA_LOCATION_ACCESS`) para criar recursos do catálogo de dados que apontem para os locais do Amazon S3. As permissões de localização de dados controlam a capacidade de criar catálogos, bancos de dados e tabelas do Data Catalog que apontam para os locais específicos do Amazon S3.
2. Para compartilhar recursos do Catálogo de Dados com outra conta no modo de acesso híbrido (sem remover as permissões de `IAMAllowedPrincipals` grupo do recurso), você precisa atualizar as configurações da versão da conta cruzada para a versão 4 ou superior. Para atualizar a versão usando o console do Lake Formation, escolha Versão 4 ou Versão 5 em Configurações de versão entre contas na página de configurações do Catálogo de Dados.

   Você também pode usar o `put-data-lake-settings` AWS CLI comando para definir o `CROSS_ACCOUNT_VERSION` parâmetro para a versão 4 ou 5:

```
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": "5"
    }
}
```

3. Para conceder permissões entre contas no modo de acesso híbrido, o concedente deve ter as permissões necessárias do IAM e os AWS Glue serviços. AWS RAM A política AWS gerenciada `AWSLakeFormationCrossAccountManager` concede as permissões necessárias.
Para permitir o compartilhamento de dados entre contas no modo de acesso híbrido, atualizamos a política gerenciada `AWSLakeFormationCrossAccountManager` adicionando duas novas permissões do IAM:

   - RAM: ListResourceSharePermissions
   - RAM: AssociateResourceSharePermission

   > **Note**
   >
   > Se você não estiver usando a política AWS gerenciada para a função de concedente, adicione as políticas acima às suas políticas personalizadas.

Localização do bucket do Amazon S3 e acesso do usuário

Ao criar um catálogo, banco de dados ou tabela no AWS Glue Data Catalog, você pode especificar a localização dos dados subjacentes no bucket do Amazon S3 e registrá-los no Lake Formation. As tabelas abaixo descrevem como as permissões funcionam para usuários (diretores) AWS Glue e usuários do Lake Formation com base na localização dos dados da tabela ou do banco de dados no Amazon S3.

Localização do Amazon S3 registrada no Lake Formation

Localização de um banco de dados no Amazon S3	AWS Glue usuários	Usuários do Lake Formation
Registrados no Lake Formation (no modo de	Tenha read/write acesso à localização de dados do Amazon S3 herdando as	Herde as permissões para criar tabelas usando a

Localização de um banco de dados no Amazon S3	AWS Glue usuários	Usuários do Lake Formation
acesso híbrido ou no modo do Lake Formation)	permissões do IAMAllowedPrincipals grupo (superacesso).	permissão CREATE TABLE concedida.
Nenhuma localização do Amazon S3 associada	Permissão DATA LOCATION explícita necessária para executar as instruções CREATE TABLE e INSERT TABLE.	Permissão DATA LOCATION explícita necessária para executar as instruções CREATE TABLE e INSERT TABLE.

IsRegisteredWithLakeFormation propriedade da tabela

A propriedade `IsRegisteredWithLakeFormation` de uma tabela indica se a localização dos dados da tabela está registrada no Lake Formation para o solicitante. Se o modo de permissão da localização estiver registrado como Lake Formation, a propriedade `IsRegisteredWithLakeFormation` será `true` para todos os usuários que acessam a localização dos dados, pois considera-se que todos os usuários optaram por essa tabela. Se a localização estiver registrada no modo de acesso híbrido, o valor será definido como `true` somente para usuários que optaram por essa tabela.

Como a **`IsRegisteredWithLakeFormation`** funciona

Modo de permissão	Users/Roles	**`IsRegisteredWithLakeFormation`**	Description
Lake Formation	Todos	Verdadeiro	Quando uma localização for registrada no Lake Formation, a propriedade de `IsRegisteredWithLakeFormation` será definida como verdadeira para

Modo de permissão	Users/Roles	**`IsRegisteredWithLakeFormation`**	Description
			todos os usuários. Isso significa que as permissões definidas no Lake Formation se aplicam à localizaç ão registrada. O fornecimento de credenciais será feito pelo Lake Formation.
Modo de acesso híbrido	Optou por usar	Verdadeiro	Com relação aos usuários que optaram por usar o Lake Formation para acesso aos dados e governança de uma tabela, a proprieda de `IsRegisteredWithLakeFormation` será definida como `true` para essa tabela. Eles estão sujeitos às políticas de permissão definidas no Lake Formation referentes à localização registrad a.

Modo de permissão	Users/Roles	**IsRegisteredWithLakeFormation**	Description
Modo de acesso híbrido	Optou por não usar	Falso	Quanto aos usuários que optaram por não usar as permissões do Lake Formation, a propriedade de `IsRegisteredWithLakeFormation` é definida como `false`. Eles não estão sujeitos às políticas de permissão definidas no Lake Formation referentes à localização registrada. Em vez disso, os usuários seguirão as políticas de permissões do Amazon S3.

Convertendo um AWS Glue recurso para um recurso híbrido

Siga estas etapas para registrar uma localização do Amazon S3 no modo de acesso híbrido e integrar novos usuários do Lake Formation sem interromper o acesso aos dados dos usuários existentes do catálogo de dados.

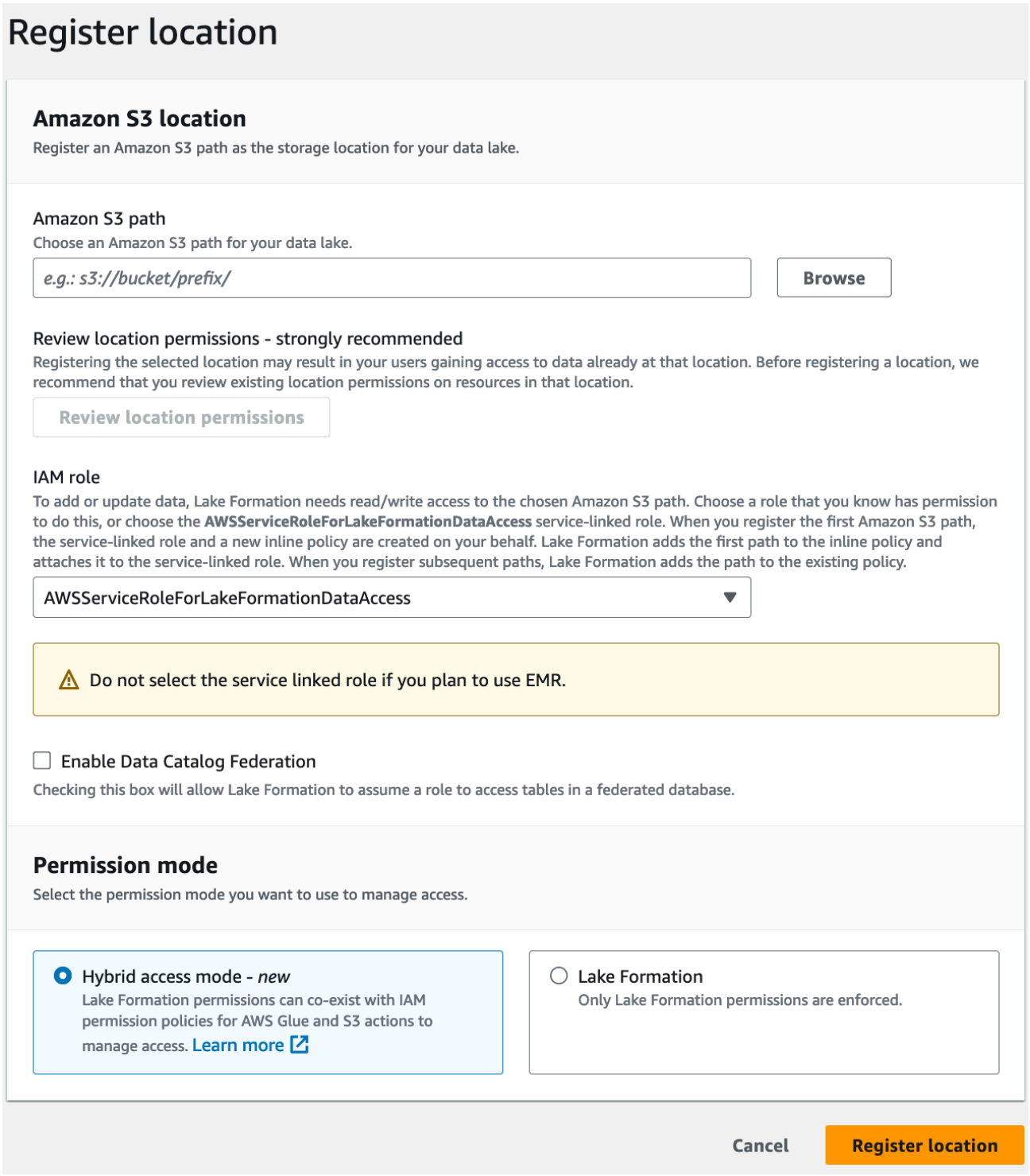
Descrição do cenário - O local dos dados não está registrado no Lake Formation, e o acesso dos usuários ao banco de dados e às tabelas do catálogo de dados é determinado pelas políticas de permissões do IAM para o Amazon S3 e ações do AWS Glue .
Por padrão, o grupo `IAMAllowedPrincipals` tem permissões `Super` em todas as tabelas do banco de dados.

Para ativar o modo de acesso híbrido para um local de dados que não está registrado no Lake Formation

1. Registre um local do Amazon S3 para ativar o modo de acesso híbrido.

   Console

   1. Faça login no console do Lake Formation como administrador do data lake.
   2. No painel de navegação, escolha Localizações do data lake em Administração.
   3. Escolha Registrar localizações.



4. Na janela Registrar localização, escolha o caminho do Amazon S3 que você deseja registrar no Lake Formation.
5. Para o perfil do IAM, escolha a função `AWSServiceRoleForLakeFormationDataAccess` vinculada ao serviço (a padrão)

ou um perfil personalizado do IAM que atenda aos requisitos em Requisitos para funções usadas para registrar locais.

6. Escolha o Modo de acesso híbrido para aplicar políticas refinadas de controle de acesso do Lake Formation às entidades principais e bancos de dados e tabelas do catálogo de dados que apontam para a localização registrada.

   Escolha Lake Formation para permitir que o Lake Formation autorize solicitações de acesso ao local registrado.

7. Escolha Registrar local.

AWS CLI

A seguir está um exemplo para registrar uma localização de dados no Lake Formation com HybridAccessEnabled:true/false. O valor padrão do parâmetro `HybridAccessEnabled` é falso. Substitua o caminho, o nome da função e o ID da AWS conta do Amazon S3 por valores válidos.

```
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. Conceda permissões e opte pelas entidades principais para usar as permissões do Lake Formation para recursos no modo de acesso híbrido

Antes de aceitar entidades principais e recursos no modo de acesso híbrido, verifique se existem permissões `Super` ou `All` para o grupo `IAMAllowedPrincipals` nos bancos de dados e tabelas que têm localização registrada no Lake Formation no modo de acesso híbrido.

> **Note**
>
> Você não pode conceder ao grupo `IAMAllowedPrincipals` permissão para `All tables` em um banco de dados. Você precisa selecionar cada tabela separadamente no menu suspenso e conceder permissões. Além disso, ao criar tabelas no banco de dados, você pode optar por usar `Use only IAM access control for new`

`tables in new databases` em Configurações do Catálogo de Dados. Essa opção concede permissão `Super` ao grupo `IAMAllowedPrincipals` automaticamente quando você cria novas tabelas no banco de dados.

Console

1. No console do Lake Formation, em Data Catalog, escolha Catálogos, Bancos de dados ou Tabelas.
2. Selecione um catálogo, um banco de dados ou uma tabela na lista e escolha Conceder no menu Ações.
3. Escolha os principais para conceder permissões no banco de dados, tabelas e colunas usando o método de recurso nomeado ou LF-Tags.

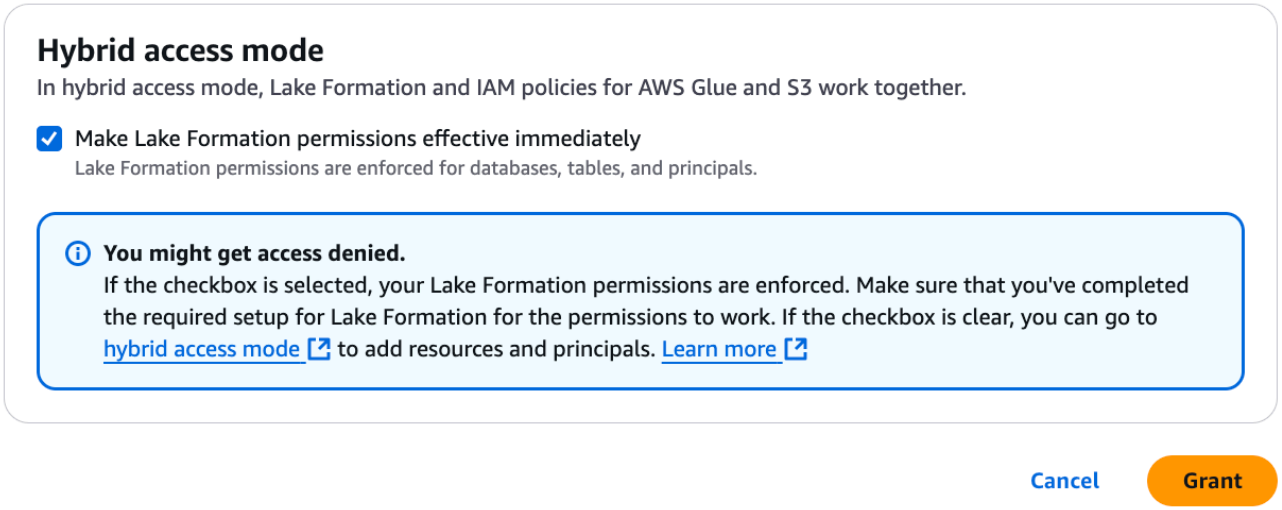
   Como alternativa, escolha Permissões de dados, selecione as entidades principais para conceder permissões na lista e escolha Conceder.

   Para obter mais detalhes sobre a concessão de permissões de dados, consulte Conceder permissões nos recursos do Catálogo de Dados.

   > **Note**
   >
   > Se você estiver concedendo a permissão Criar tabela a uma entidade principal, também precisará conceder permissões de localização de dados (`DATA_LOCATION_ACCESS`) à entidade principal. Essa permissão não é necessária para atualizar tabelas.
   > Para obter mais informações, consulte Conceder permissões de localização de dados.

4. Quando você usa o Método de recurso nomeado para conceder permissões, a opção de optar por entidades principais e recursos está disponível na seção inferior da página Conceder permissão de dados.

   Escolha Tornar as permissões do Lake Formation efetivas imediatamente para habilitar as permissões do Lake Formation para as entidades principais e recursos.



5. Selecione Conceder.

   Quando você opta pela entidade principal A na tabela A que está apontando para um local de dados, ela permite que a entidade principal A tenha acesso ao local dessa tabela usando as permissões do Lake Formation se o local dos dados estiver registrado no modo híbrido.

AWS CLI

A seguir está um exemplo de como optar por uma entidade principal e uma tabela no modo de acesso híbrido. Substitua o nome da função, o ID da conta da AWS , o nome do banco de dados e o nome da tabela por valores válidos.

```
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>"
            }
        }
```

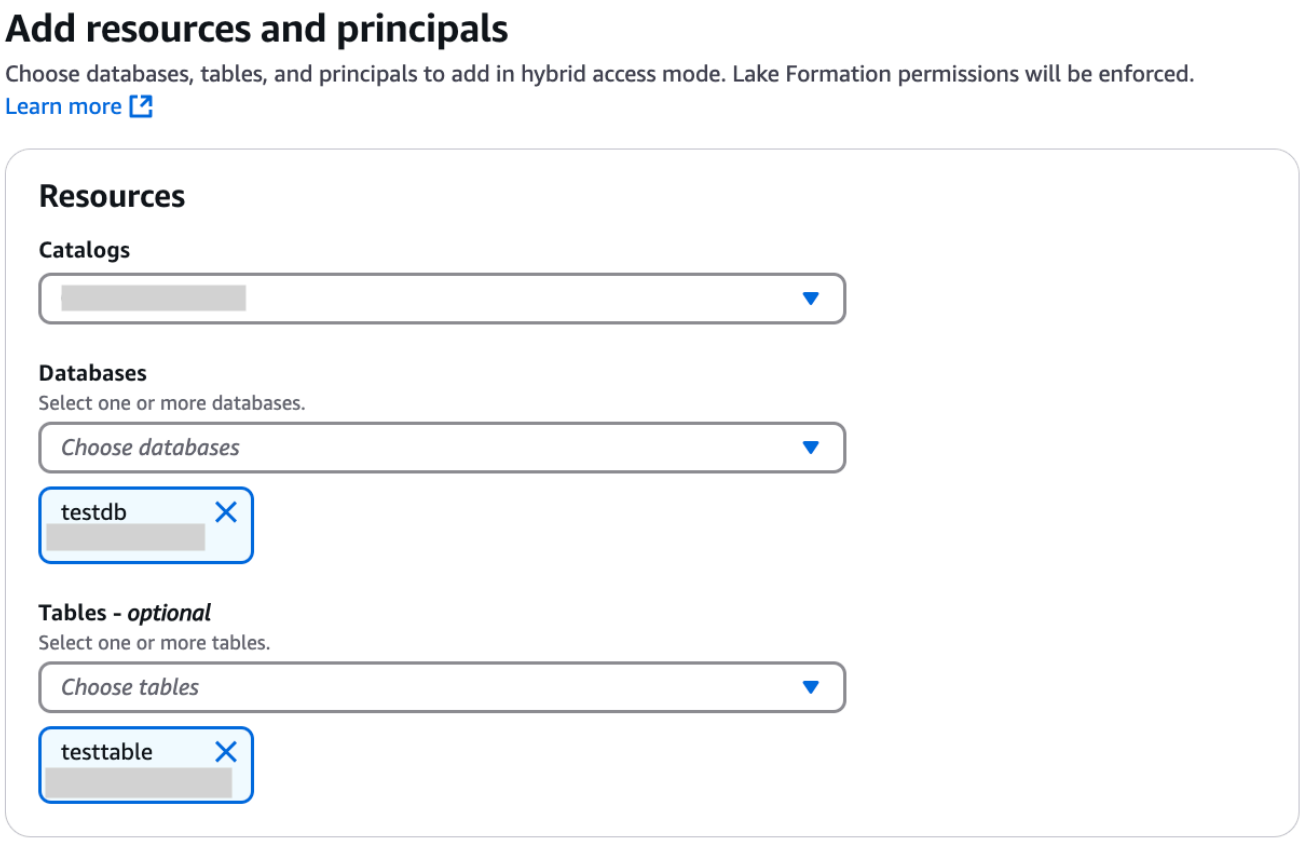
```
    }
```

a. (Optional) Se você optar por LF-Tags conceder permissões, poderá optar que os diretores usem as permissões do Lake Formation em uma etapa separada. Você pode fazer isso escolhendo o Modo de acesso híbrido em Permissões na barra de navegação esquerda.

b. Na seção inferior da página do Modo de acesso híbrido, escolha Adicionar para adicionar recursos e entidades principais ao modo de acesso híbrido.

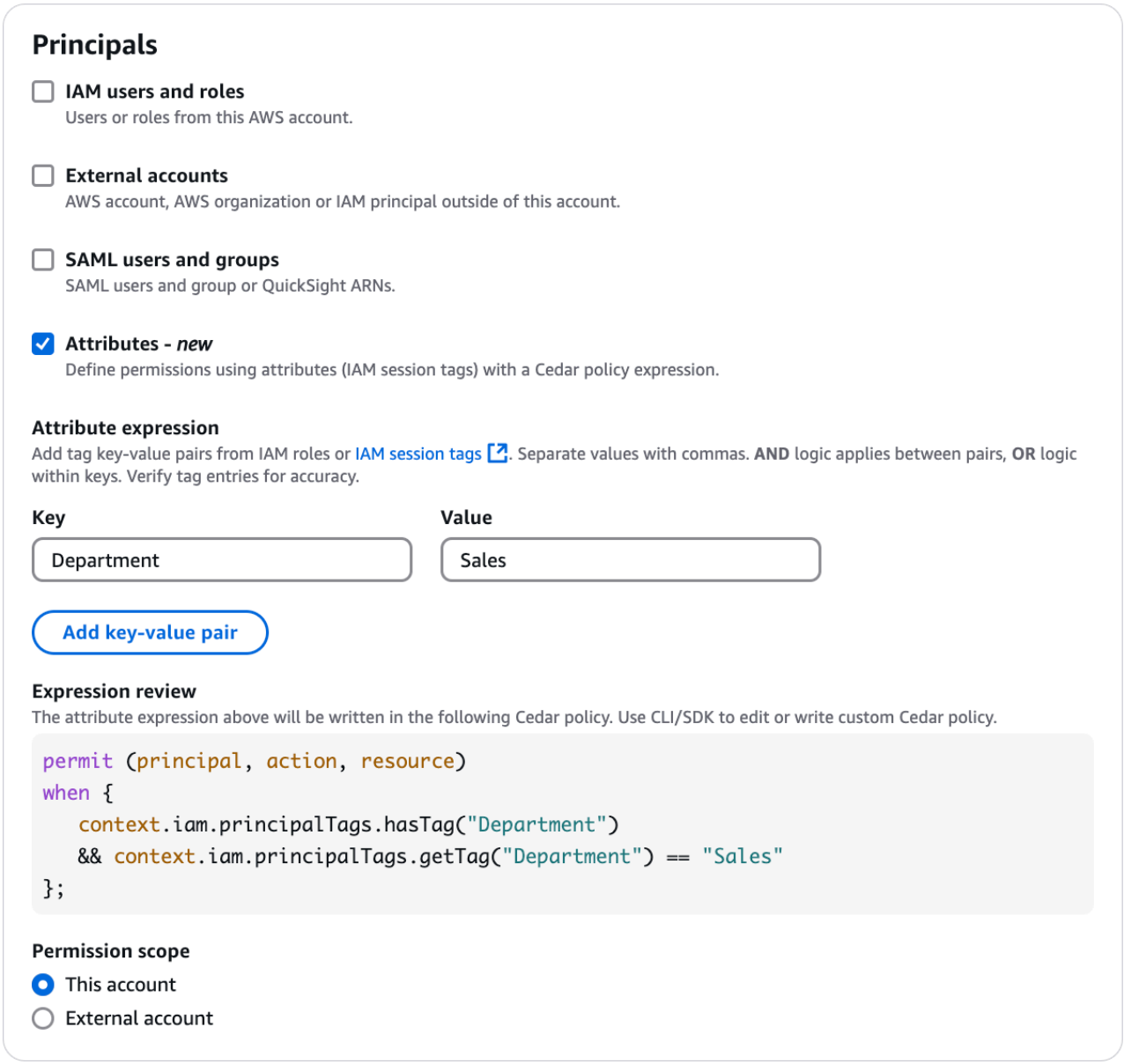
c. Na página Adicionar recursos e entidades principais, escolha os catálogos, os bancos de dados e as tabelas registrados no modo de acesso híbrido.

   Você pode escolher `All tables` em um banco de dados para conceder acesso.



d. Escolha entidades principais para aceitarem o uso das permissões do Lake Formation no modo de acesso híbrido.

   - Entidade principais: é possível escolher usuários e perfis do IAM na mesma conta ou em outra conta. Você também pode escolher usuários e grupos SAML.
   - Atributos: selecione atributos para conceder permissões com base em atributos.



- Insira o par de chave-valor para criar uma concessão com base em atributos. Analise a expressão da política Cedar no console. Para acessar mais informações sobre Cedar, consulte O que é Cedar? | Referência GuideLink da linguagem Cedar Policy.
- Escolha Adicionar.

  Todos os IAM roles/users com atributos correspondentes recebem acesso.

e. Escolha Adicionar.

Convertendo um recurso do Lake Formation em um recurso híbrido

Nos casos em que você está usando atualmente as permissões do Lake Formation para seus bancos de dados e tabelas do catálogo de dados, você pode editar as propriedades de registro de localização para ativar o modo de acesso híbrido. Isso permite que você forneça aos novos diretores acesso aos mesmos recursos usando políticas de permissão do IAM para o Amazon S3 AWS Glue e ações sem interromper as permissões existentes do Lake Formation.

Descrição do cenário - As etapas a seguir pressupõem que você tenha um local de dados registrado no Lake Formation e tenha configurado permissões para entidades principais em bancos de dados, tabelas ou colunas que apontam para esse local. Se o local foi registrado com uma função vinculada ao serviço, você não poderá atualizar os parâmetros de localização e ativar o modo de acesso híbrido. Por padrão, o grupo `IAMAllowedPrincipals` tem permissões Super no banco de dados e em todas as respectivas tabelas.

> **Important**
>
> Não atualize um registro de localização para o modo de acesso híbrido sem optar pelas entidades principais que estão acessando os dados nesse local.

Ativando o modo de acesso híbrido para um local de dados registrado no Lake Formation

1.
> **Warning**
>
> Não recomendamos converter uma localização de dados gerenciada do Lake Formation no modo de acesso híbrido para evitar a interrupção das políticas de permissões de outros usuários ou workloads existentes.

   Opte pelas entidades principais existentes que tenham permissões do Lake Formation.

   1. Liste e analise as permissões que você concedeu às entidades principais em catálogos, bancos de dados e tabelas. Para obter mais informações, consulte Visualizar permissões de banco de dados e tabelas no Lake Formation.
   2. Escolha o Modo de acesso híbrido em Permissões na barra de navegação esquerda e escolha Adicionar.

3. Na página Adicionar entidades principais e recursos, escolha os catálogos, os bancos de dados e as tabelas do local de dados do Amazon S3 que você deseja usar no modo de acesso híbrido. Escolha as entidades principais que já possuem permissões do Lake Formation.
4. Escolha Adicionar para ativar as entidades principais para usar as permissões do Lake Formation no modo de acesso híbrido.

2. Atualize o bucket/prefix registro do Amazon S3 escolhendo a opção de modo de acesso híbrido.

   Console

   1. Faça login no console do Lake Formation como administrador do data lake.
   2. No painel de navegação, em Registrar e ingerir, escolha Locais do data lake.
   3. Selecione um local e, no menu Ações, escolha Editar.
   4. Escolha o Modo de acesso híbrido.
   5. Escolha Salvar.
   6. Em catálogo de dados, selecione o banco de dados ou a tabela e conceda permissões `Super` ou `All` para o grupo virtual chamado `IAMAllowedPrincipals`.
   7. Verifique se o acesso dos usuários existentes do Lake Formation não foi interrompido quando você atualizou as propriedades de registro de localização. Faça login no console do Athena como entidade principal do Lake Formation e execute um exemplo de consulta em uma tabela que está apontando para o local atualizado.

      Da mesma forma, verifique o acesso dos AWS Glue usuários que estão usando as políticas de permissões do IAM para acessar o banco de dados e as tabelas.

   AWS CLI

   A seguir está um exemplo para registrar uma localização de dados no Lake Formation com HybridAccessEnabled:true/false. O valor padrão do parâmetro `HybridAccessEnabled` é falso. Substitua o caminho, o nome da função e o ID da AWS conta do Amazon S3 por valores válidos.

```
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
}
```

Compartilhando um AWS Glue recurso usando o modo de acesso híbrido

Compartilhe dados com outra pessoa Conta da AWS ou com um diretor de outra pessoa Conta da AWS aplicando as permissões do Lake Formation sem interromper o acesso baseado em IAM dos usuários existentes do Catálogo de Dados.

Descrição do cenário - A conta do produtor tem um banco de dados do catálogo de dados que tem acesso controlado usando as principais políticas do IAM para Amazon S3 e AWS Glue ações. A localização dos dados do banco de dados não está registrada no Lake Formation. Por padrão, o grupo `IAMAllowedPrincipals` tem permissões `Super` no banco de dados e em todas as respectivas tabelas.

Conceder permissões entre contas do Lake Formation no modo de acesso híbrido

1. Configuração da conta de produtor

    1. Faça login no console do Lake Formation usando uma função que tenha permissão do IAM `lakeformation:PutDataLakeSettings`.
    2. Acesse as Configurações do catálogo de dados e escolha `Version 4` para as Configurações da versão entre contas.

       Se você estiver usando a versão 1 ou 2, consulte as instruções Como atualizar as configurações da versão de compartilhamento de dados entre contas sobre como atualizar para a versão 3.

       Não são necessárias alterações na política de permissão ao atualizar da versão 3 para a 4.
    3. Registre a localização do Amazon S3 do banco de dados ou tabela que você planeja compartilhar no modo de acesso híbrido.
    4. Verifique se a permissão `Super` para o grupo `IAMAllowedPrincipals` existe nos bancos de dados e tabelas nos quais você registrou a localização dos dados no modo de acesso híbrido na etapa acima.
    5. Conceda permissões do Lake Formation para AWS organizações, unidades organizacionais (OUs) ou diretamente com um diretor do IAM em outra conta.

6. Se você estiver concedendo permissões diretamente a uma entidade principal do IAM, opte pela entidade principal da conta de consumidor para aplicar as permissões do Lake Formation no modo de acesso híbrido, ativando a opção Tornar as permissões do Lake Formation efetivas imediatamente.

   Se você estiver concedendo permissões entre contas para outra AWS conta, ao optar pela conta, as permissões do Lake Formation serão aplicadas somente para os administradores dessa conta. O administrador do data lake da conta do destinatário precisa reduzir as permissões em cascata e optar pelas entidades principais da conta para aplicar as permissões do Lake Formation aos recursos compartilhados que estão no modo de acesso híbrido.

   Se você escolher a LF-Tags opção Recursos combinados por para conceder permissões entre contas, primeiro precisará concluir a etapa de concessão de permissões. Você pode optar por incluir entidades principais e recursos no modo de acesso híbrido como uma etapa separada, escolhendo o Modo de acesso híbrido em Permissões na barra de navegação esquerda do console do Lake Formation. Em seguida, escolha Adicionar para adicionar os recursos e as entidades principais aos quais você deseja aplicar as permissões do Lake Formation.

2. Configuração de conta de consumidor

   1. Faça login no console do Lake Formation https://console.aws.amazon.com/lakeformation/como administrador do data lake.
   2. Acesse https://console.aws.amazon.com/ram/homee aceite o convite de compartilhamento de recursos. A guia Compartilhado comigo no AWS RAM console exibe o banco de dados e as tabelas compartilhadas com sua conta.
   3. Crie um link de recurso para a and/or tabela do banco de dados compartilhado no Lake Formation.
   4. Conceda a permissão `Describe` no link do recurso e a permissão `Grant on target` (no recurso compartilhado original) às entidades principais do IAM em sua conta (de consumidor).
   5. Conceda permissões do Lake Formation no banco de dados ou na tabela compartilhada com você às entidades principais da sua conta. Opte pelas entidades principais e recursos para aplicar as permissões do Lake Formation no modo de acesso híbrido, ativando a opção Tornar as permissões do Lake Formation efetivas imediatamente.

6. Teste as permissões da entidade principal do Lake Formation executando exemplos de consultas do Athena. Teste o acesso existente de seus AWS Glue usuários com as principais políticas do IAM para Amazon S3 e AWS Glue ações.

   (Opcional) Remova a política de bucket do Amazon S3 para acesso a dados e políticas de entidades principais do IAM para AWS Glue e acesso a dados do Amazon S3 para as entidades principais que você configurou para usar permissões do Lake Formation.

Compartilhando um recurso do Lake Formation usando o modo de acesso híbrido

Permita que novos usuários do catálogo de dados em uma conta externa acessem bancos de dados e tabelas do catálogo de dados usando políticas baseadas no IAM sem interromper as permissões de compartilhamento entre contas existentes do Lake Formation.

Descrição do cenário - A conta de produtor tem banco de dados e tabelas gerenciados pelo Lake Formation que são compartilhados com uma conta externa (consumidor) no nível da conta ou no nível de entidade principal do IAM. A localização dos dados do banco de dados é registrada no Lake Formation. O grupo `IAMAllowedPrincipals` não tem permissões `Super` no banco de dados e em suas tabelas.

Conceder acesso entre contas a novos usuários do catálogo de dados por meio de políticas baseadas em IAM sem interromper as permissões existentes do Lake Formation

1. Configuração da conta de produtor

   1. Faça login no console do Lake Formation usando uma função que `lakeformation:PutDataLakeSettings`.
   2. Em Configurações do catálogo de dados, escolha `Version 4` para as Configurações da versão entre contas.

      Se você estiver usando a versão 1 ou 2, consulte as instruções Como atualizar as configurações da versão de compartilhamento de dados entre contas sobre como atualizar para a versão 3.

      Não são necessárias alterações na política de permissão para atualizar da versão 3 para a 4.
   3. Liste as permissões que você concedeu às entidades principais em bancos de dados e tabelas. Para obter mais informações, consulte Visualizar permissões de banco de dados e tabelas no Lake Formation.

4. Conceda novamente as permissões existentes entre contas do Lake Formation optando por entidades principais e recursos.

> **Note**
>
> Antes de atualizar um registro de localização de dados para o modo de acesso híbrido para conceder permissões entre contas, você precisa conceder novamente pelo menos um compartilhamento de dados entre contas por conta. Essa etapa é necessária para atualizar as permissões AWS RAM gerenciadas anexadas ao compartilhamento AWS RAM de recursos.
> Em julho de 2023, o Lake Formation atualizou as permissões AWS RAM gerenciadas usadas para compartilhar bancos de dados e tabelas:
>
> - `arn:aws:ram::aws:permission/AWSRAMLFEnabledGlueAllTablesReadWriteForDatabase` (política de compartilhamento em nível de banco de dados)
> - `arn:aws:ram::aws:permission/AWSRAMLFEnabledGlueTableReadWrite` (política de compartilhamento em nível de tabela)
>
> As concessões de permissão entre contas feitas antes de julho de 2023 não têm essas AWS RAM permissões atualizadas.
> Se você concedeu permissões entre contas diretamente às entidades principais, precisará devolvê-las individualmente às entidades principais. Se você pular essa etapa, as entidades principais que acessam o recurso compartilhado podem receber um erro de combinação ilegal.

5. Acesse https://console.aws.amazon.com/ram/home.
6. A guia Compartilhado por mim no AWS RAM console exibe os nomes do banco de dados e da tabela que você compartilhou com uma conta externa ou principal.

   Certifique-se de que as permissões anexadas ao recurso compartilhado tenham o ARN correto.
7. Verifique se os recursos no AWS RAM compartilhamento estão no `Associated` status. Se o status for exibido como `Associating`, espere até que eles entrem no status `Associated`. Se o status for `Failed`, pare e entre em contato com a equipe de serviço do Lake Formation.
8. Escolha o Modo de acesso híbrido em Permissões na barra de navegação esquerda e escolha Adicionar.

9. A página Adicionar diretores e recursos mostra os bancos de dados, and/or as tabelas e os principais que têm acesso. Você pode fazer as atualizações necessárias adicionando ou removendo entidades principais e recursos.
10. Escolha as entidades principais com permissões do Lake Formation para o banco de dados e as tabelas que você deseja alterar para o modo de acesso híbrido. Escolha os bancos de dados e tabelas.
11. Escolha Adicionar para optar pelas entidades principais para aplicar as permissões do Lake Formation no modo de acesso híbrido.
12. Conceda a permissão `Super` ao grupo virtual `IAMAllowedPrincipals` em seu banco de dados e nas tabelas selecionadas.
13. Edite o registro Lake Formation da localização do Amazon S3 para o modo de acesso híbrido.
14. Conceda permissões para os AWS Glue usuários na conta externa (consumidor) usando políticas de permissão do IAM para ações do Amazon S3 AWS Glue .

2. Configuração de conta de consumidor

   1. Faça login no console do Lake Formation https://console.aws.amazon.com/lakeformation/como administrador do data lake.
   2. Acesse https://console.aws.amazon.com/ram/homee aceite o convite de compartilhamento de recursos. A guia Recursos compartilhados comigo na AWS RAM página exibe os nomes do banco de dados e das tabelas que são compartilhados com sua conta.

      Para o AWS RAM compartilhamento, certifique-se de que a permissão anexada tenha o ARN correto do convite compartilhado AWS RAM . Verifique se os recursos no AWS RAM compartilhamento estão no `Associated` status. Se o status for exibido como `Associating`, espere até que eles entrem no status `Associated`. Se o status for `Failed`, pare e entre em contato com a equipe de serviço do Lake Formation.
   3. Crie um link de recurso para a and/or tabela do banco de dados compartilhado no Lake Formation.
   4. Conceda a permissão `Describe` no link do recurso e a permissão `Grant on target` (no recurso compartilhado original) às entidades principais do IAM em sua conta (de consumidor).
   5. Em seguida, configure as permissões do Lake Formation para as entidades principais da sua conta no banco de dados ou na tabela compartilhada.

      Na barra de navegação esquerda, em Permissões, escolha Modo de acesso híbrido.

6. Escolha Adicionar na seção inferior da página do Modo de acesso híbrido para optar pelas entidades principais e pelo banco de dados ou tabela compartilhados com você na conta de produtor.
7. Conceda permissões para os AWS Glue usuários em sua conta usando políticas de permissão do IAM para ações do Amazon S3 AWS Glue .
8. Teste as permissões e AWS Glue permissões do Lake Formation dos usuários executando exemplos de consultas separadas na tabela usando o Athena

   (Opcional) Limpe as políticas de permissão do IAM para o Amazon S3 para as entidades principais que estão no modo de acesso híbrido.

Removendo entidades principais e recursos do modo de acesso híbrido

Siga estas etapas para remover bancos de dados, tabelas e entidades principais do modo de acesso híbrido.

Console

1. Faça login no console do Lake Formation em https://console.aws.amazon.com/lakeformation/.
2. Em Permissões, escolha o Modo de acesso híbrido.
3. Na página do Modo de acesso híbrido, marque a caixa de seleção ao lado do nome do banco de dados ou da tabela e escolha `Remove`.
4. Uma mensagem de aviso solicita que você confirme a ação. Escolha Remover .

   O Lake Formation não impõe mais permissões para esses recursos, e o acesso a esse recurso será controlado usando IAM e AWS Glue permissões. Isso pode fazer com que o usuário não tenha mais acesso a esse recurso se não tiver as permissões apropriadas do IAM.

AWS CLI

O exemplo a seguir mostra como remover um recurso do modo de acesso híbrido.

```
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>"
          }
    }
}
```

Visualizando entidades principais e recursos no modo de acesso híbrido

Siga estas etapas para visualizar bancos de dados, tabelas e entidades principais no modo de acesso híbrido.

Console

1. Faça login no console do Lake Formation em https://console.aws.amazon.com/lakeformation/.
2. Em Permissões, escolha o Modo de acesso híbrido.
3. A página do Modo de acesso híbrido mostra os recursos e entidades principais que estão atualmente no modo de acesso híbrido.

AWS CLI

O exemplo a seguir mostra como listar todas as entidades principais e recursos opcionais que estão no modo de acesso híbrido.

```
aws lakeformation list-lake-formation-opt-ins
```

O exemplo a seguir mostra como listar o opt-in para um par específico de entidade principal e recurso.

```
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>"
          }
    }
}
```

Recursos adicionais do

Na postagem do blog a seguir, mostraremos as instruções para integrar as permissões do Lake Formation no modo de acesso híbrido para usuários selecionados, enquanto o banco de dados já está acessível a outros usuários por meio das permissões do IAM e do Amazon S3. Analisaremos as instruções para configurar o modo de acesso híbrido em uma AWS conta e entre duas contas.

- Apresentando o modo de acesso híbrido AWS Glue Data Catalog para proteger o acesso usando as políticas Lake Formation e IAM e Amazon S3.

Criação de objetos no AWS Glue Data Catalog

AWS Lake Formation usa o AWS Glue Data Catalog (Catálogo de dados) para armazenar metadados sobre lagos de dados, fontes de dados, transformações e destinos. Metadados são dados sobre os dados subjacentes em seu conjunto de dados. Cada AWS conta tem um catálogo de dados por AWS região.

Os metadados no Data Catalog são organizados em uma hierarquia de dados de três níveis que inclui catálogos, bancos de dados e tabelas. Ele organiza dados de várias fontes em contêineres lógicos chamados catálogos. Cada catálogo representa dados de fontes, como data warehouses do Amazon Redshift, bancos de dados do Amazon DynamoDB e fontes de dados de terceiros, como Snowflake, MySQL, e mais de trinta fontes de dados externas, que são integradas por meio de conectores federados. Você também pode criar catálogos no Data Catalog para armazenar dados em buckets de Tabelas do S3 ou no Redshift Managed Storage (RMS).

As tabelas armazenam informações sobre os dados subjacentes, incluindo informações de esquema, informações de partição e localização dos dados. Bancos de dados são coleções de tabelas. O Data Catalog também contém links de recursos, que são links para catálogos, bancos de dados e tabelas compartilhados em contas externas e são usados para acesso entre contas aos dados no data lake.

O Data Catalog é um objeto de catálogo aninhado que contém catálogos, bancos de dados e tabelas. Ele é referenciado pelo Conta da AWS ID e é o catálogo padrão em uma conta e em uma Região da AWS. O Data Catalog usa uma hierarquia de três níveis (catalog.database.table) para organizar tabelas.

- Catálogo: o nível mais alto da hierarquia de metadados de três níveis do Data Catalog. Você pode adicionar vários catálogos no Data Catalog por meio da federação.
- Banco de dados: o segundo nível da hierarquia de metadados, composto por tabelas e visualizações. Um banco de dados também é chamado de esquema em muitos sistemas de dados, como Amazon Redshift e Trino.
- Tabela e visualização: o terceiro nível da hierarquia de dados de três níveis do Data Catalog.

Todas as tabelas Iceberg no Amazon S3 são armazenadas no catálogo de dados padrão com ID do catálogo = ID Conta da AWS . Você pode criar catálogos federados para armazenar definições de tabelas no AWS Glue Data Catalog Amazon Redshift, no armazenamento de tabelas do Amazon S3 ou em outras fontes de dados de terceiros por meio da federação.

Tópicos

- Criar um catálogo
- Criação de um banco de dados
- Criar tabelas
- Desenvolvimento AWS Glue Data Catalog visualizações

Criar um catálogo

Os catálogos representam o nível mais alto ou mais alto na hierarquia de metadados de três níveis do AWS Glue Data Catalog. Você pode usar vários métodos para trazer dados para o Data Catalog e criar catálogos de vários níveis.

Para acessar mais informações sobre a criação de catálogos por meio de fontes de dados externas, consulte Trazendo seus dados para o AWS Glue Data Catalog.

Para criar um catálogo usando o console do Lake Formation, você deve estar conectado como administrador do data lake ou criador do catálogo. Um criador de catálogo é uma entidade principal que recebeu a permissão `CREATE_CATALOG` do Lake Formation. Você pode ver uma lista de criadores de catálogos na página Perfis e tarefas administrativas do console do Lake Formation. Para ver essa lista, você precisa ter a permissão `lakeformation:ListPermissions` do IAM e estar conectado como administrador do data lake ou como criador de catálogos com a opção de concessão na permissão `CREATE_CATALOG`.

Criação de um banco de dados

As tabelas de metadados no catálogo de dados são armazenadas nos bancos de dados. Você pode criar quantos bancos de dados precisar e conceder permissões diferentes do Lake Formation em cada banco de dados.

Os bancos de dados podem ter uma propriedade de localização opcional. Esse local geralmente está dentro de um local do Amazon Simple Storage Service (Amazon S3) registrado no Lake Formation. Quando você especifica um local, as entidades principais não precisam de permissões de localização de dados para criar tabelas do catálogo de dados que apontem para locais dentro do local do banco de dados. Para obter mais informações, consulte Underlying data access control.

Para criar um banco de dados usando o console do Lake Formation, você deve estar conectado como administrador do data lake ou criador do banco de dados. Um criador de banco de dados é uma entidade principal que recebeu a permissão `CREATE_DATABASE` do Lake Formation. Você pode ver uma lista de criadores de banco de dados na página Funções e tarefas administrativas do console do Lake Formation. Para ver essa lista, você precisa ter a permissão `lakeformation:ListPermissions` do IAM e estar conectado como administrador do data lake ou como criador de banco de dados com a opção de concessão na permissão `CREATE_DATABASE`.

Para criar um banco de dados

1. Abra o AWS Lake Formation console em https://console.aws.amazon.com/lakeformation/e faça login como administrador do data lake ou criador do banco de dados.
2. No painel de navegação, em catálogo de dados, escolha Bancos de dados.
3. Selecione Criar banco de dados.
4. Na caixa de diálogo Criar banco de dados, insira o nome do banco de dados, a localização opcional e a descrição opcional.
5. Opcionalmente, selecione Usar somente controle de acesso do IAM para novas tabelas nesse banco de dados.

Para obter mais informações sobre esta opção, consulte the section called “Alterando as configurações padrão do seu data lake”.

6. Selecione Criar banco de dados.

Criar tabelas

AWS Lake Formation as tabelas de metadados contêm informações sobre dados no data lake, incluindo informações de esquema, informações de partição e localização dos dados. Essas tabelas são armazenadas no catálogo de dados do AWS Glue. Você os usa para acessar dados subjacentes no data lake e gerenciar esses dados com as permissões do Lake Formation. As tabelas são armazenadas nos bancos de dados no catálogo de dados.

Há várias maneiras de criar tabelas do catálogo de dados:

- Execute um crawler no AWS Glue. Consulte Definição de crawlers no Guia do desenvolvedor do AWS Glue .
- Crie e execute um fluxo de trabalho. Consulte the section called “Importação de dados usando fluxos de trabalho”.
- Crie uma tabela manualmente usando o console do Lake Formation, a API AWS Glue ou a AWS Command Line Interface (AWS CLI).
- Crie uma tabela usando Amazon Athena.
- Crie um link de recurso para uma tabela em uma conta externa. Consulte the section called “Criação de links de recursos”.

Criar tabelas no Apache Iceberg

AWS Lake Formation suporta a criação de tabelas Apache Iceberg que usam o formato de dados Apache Parquet AWS Glue Data Catalog com dados residentes no Amazon S3. Uma tabela no catálogo de dados é a definição de metadados que representa os dados em um armazenamento de dados. Por padrão, o Lake Formation cria tabelas do Iceberg v2. Para saber a diferença entre as tabelas da v1 e v2, consulte Alterações de versão do formato na documentação do Apache Iceberg.

Apache Iceberg é um formato de tabela aberta para conjuntos de dados analíticos muito grandes. O Iceberg permite mudanças fáceis em seu esquema, também conhecido como evolução do esquema, o que significa que os usuários podem adicionar, renomear ou remover colunas de uma tabela de

dados sem interromper os dados subjacentes. O Iceberg também fornece suporte para controle de versão de dados, o que permite que os usuários acompanhem as alterações nos dados ao longo do tempo. Isso ativa o atributo de viagem no tempo, que permite que os usuários acessem e consultem versões históricas dos dados e analisem as alterações nos dados entre atualizações e exclusões.

Você pode usar o console do Lake Formation ou a `CreateTable` operação na AWS Glue API para criar uma tabela Iceberg no Catálogo de Dados. Para obter mais informações, consulte CreateTable action (Python: create_table).

Ao criar uma tabela do Iceberg no catálogo de dados, você deve especificar o formato da tabela e o caminho do arquivo de metadados no Amazon S3 para poder realizar leituras e gravações.

Você pode usar o Lake Formation para proteger sua tabela Iceberg usando permissões de controle de acesso refinadas ao registrar a localização de dados do Amazon S3 com. AWS Lake Formation Para dados de origem no Amazon S3 e metadados que não estão registrados no Lake Formation, o acesso é determinado pelas políticas de permissões do IAM para o Amazon S3 e ações. AWS Glue Para obter mais informações, consulte Gerenciando permissões do Lake Formation.

> **Note**
>
> O catálogo de dados não oferece suporte à criação de partições e à adição de propriedades da tabela do Iceberg.

Tópicos

- Pré-requisitos
- Criar uma tabela no Iceberg

Pré-requisitos

Para criar tabelas Iceberg no catálogo de dados e configurar as permissões de acesso aos dados do Lake Formation, você precisa preencher os seguintes requisitos:

1. Permissões necessárias para criar tabelas do Iceberg sem os dados registrados no Lake Formation.

   Além das permissões necessárias para criar uma tabela no catálogo de dados, o criador da tabela precisa as seguintes permissões:

- `s3:PutObject` no recurso arn:aws:s3:::{bucketName}
- `s3:GetObject` no recurso arn:aws:s3:::{bucketName}
- `s3:DeleteObject` no recurso arn:aws:s3:::{bucketName}

2. Permissões necessárias para criar tabelas do Iceberg com dados registrados no Lake Formation:

   Para usar o Lake Formation para gerenciar e proteger os dados em seu data lake, registre sua localização no Amazon S3 que tenha os dados para tabelas com o Lake Formation. Isso é para que a Lake Formation possa fornecer credenciais para serviços AWS analíticos como Athena, Redshift Spectrum e Amazon EMR para acessar dados. Para obter mais informações sobre como registrar um local do Amazon S3, consulte Adicionar uma localização do Amazon S3 ao seu data lake.

   Uma entidade principal que lê e grava os dados subjacentes registrados no Lake Formation exige as seguintes permissões:

   - `lakeformation:GetDataAccess`
   - `DATA_LOCATION_ACCESS`

     Uma entidade principal que tem permissões de localização de dados em um local também tem permissões de localização em todos os locais secundários.

     Para obter mais informações sobre permissões de localização de dados, consulte Controle de acesso a dados subjacente.

Para permitir a compactação, o serviço precisa assumir um perfil do IAM que tenha permissões para atualizar tabelas no catálogo de dados. Para obter detalhes, consulte Table optimization prerequisites.

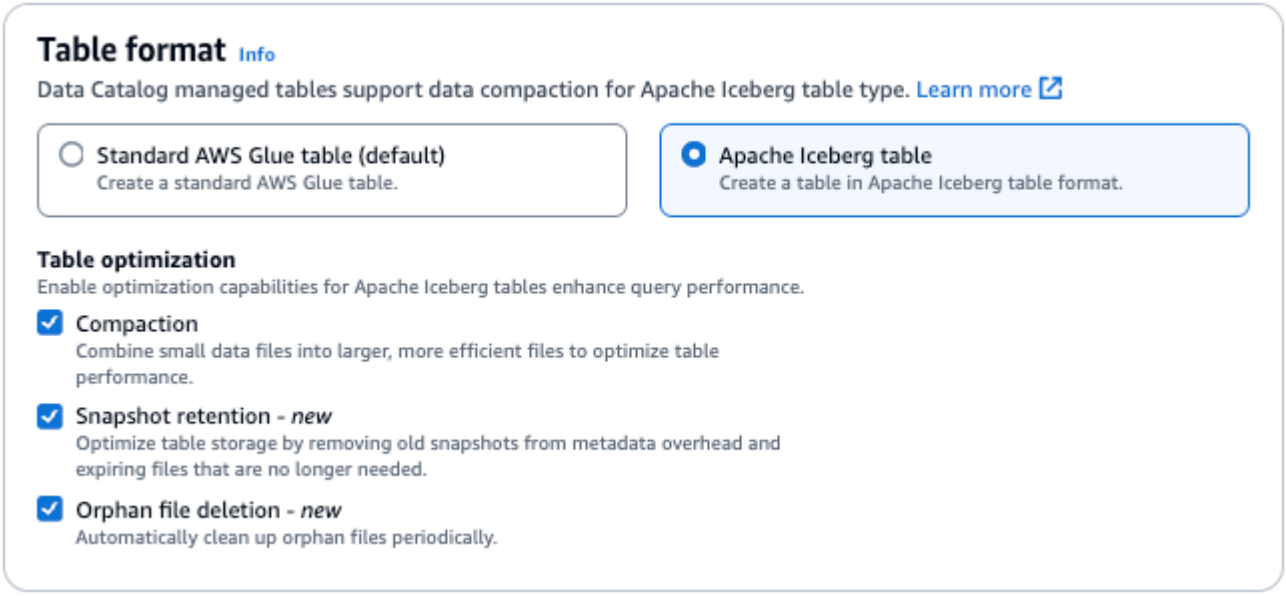
Criar uma tabela no Iceberg

Você pode criar tabelas Iceberg v1 e v2 usando o console Lake Formation ou AWS Command Line Interface conforme documentado nesta página. Você também pode criar tabelas Iceberg usando o AWS Glue console ou Crawler do AWS Glue. Para obter mais informações, consulte Catálogo de dados e crawlers no Guia do desenvolvedor do AWS Glue .

Para criar uma tabela no Iceberg

Console

1. Faça login no Console de gerenciamento da AWS, e abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.
2. Em catálogo de dados, escolha Tabelas e use o botão Criar tabela para especificar os seguintes atributos:
   - Nome da tabela: insira um nome para a tabela. Se você estiver usando o Athena para acessar tabelas, use essas dicas de nomenclatura no Guia do usuário do Amazon Athena.
   - Banco de dados: escolha um banco de dados existente ou crie um novo.
   - Descrição: descrição da tabela. Você pode escrever uma descrição para ajudá-lo a entender o conteúdo da tabela.
   - Formato da tabela: para Formato da tabela, escolha Apache Iceberg.



   - Otimização de tabelas
     - Compactação: os arquivos de dados são mesclados e regravados para remover dados obsoletos e consolidar dados fragmentados em arquivos maiores e mais eficientes.
     - Retenção de snapshots: os snapshots são versões com carimbo de data e hora de uma tabela do Iceberg. As configurações de retenção de snapshots permitem que os clientes determinem por quanto tempo reter e quantos snapshots devem ser retidos. A configuração de um otimizador de retenção de snapshots pode ajudar a gerenciar a sobrecarga de armazenamento removendo snapshots antigos e desnecessários e seus arquivos subjacentes.

- Exclusão de arquivos órfãos: arquivos órfãos são arquivos que não são mais referidos pelos metadados da tabela do Iceberg. Esses arquivos podem se acumular ao longo do tempo, especialmente após operações como exclusões de tabelas ou trabalhos de ETL com falha. A ativação da exclusão de arquivos órfãos permite identificar e AWS Glue remover periodicamente esses arquivos desnecessários, liberando espaço de armazenamento.

Para obter mais informações, consulte Como otimizar tabelas do Iceberg.

- Perfil do IAM: para executar a compactação, o serviço assume um perfil do IAM em seu nome. Você pode escolher um perfil do IAM usando o menu suspenso. Certifique-se de que a função tenha as permissões necessárias para habilitar a compactação.

  Para saber mais sobre as permissões necessárias, consulte Table optimization prerequisites.
- Localização: especifique o caminho para a pasta no Amazon S3 que armazena a tabela de metadados. O Iceberg precisa de um arquivo de metadados e de um local no catálogo de dados para poder realizar leituras e gravações.
- Esquema: escolha Adicionar colunas para adicionar colunas e tipos de dados das colunas. Você tem a opção de criar uma tabela vazia e atualizar o esquema posteriormente.
  O catálogo de dados oferece suporte aos tipos de dados do Hive. Para obter mais informações, consulte Tipos de dados do Hive.

  O Iceberg permite que você evolua o esquema e a partição depois de criar a tabela. Você pode usar as consultas do Athena para atualizar o esquema da tabela e as consultas do Spark para atualizar as partições.

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/"
            }
        }'
```

Otimizar tabelas Iceberg

O Lake Formation suporta várias opções de otimização de tabelas para aprimorar o gerenciamento e o desempenho das tabelas Apache Iceberg usadas pelos mecanismos AWS analíticos e pelas tarefas de ETL. Esses otimizadores fornecem utilização eficiente do espaço em disco, melhor performance de consultas e gerenciamento de dados. Existem três tipos de otimizador de tabelas disponíveis no Lake Formation:

- Compactação: a compactação de dados compacta pequenos arquivos de dados para reduzir o uso de armazenamento e melhorar a performance de leitura. Os arquivos de dados são mesclados e regravados para remover dados obsoletos e consolidar dados fragmentados em arquivos maiores e mais eficientes. A compactação pode ser configurada para ser executada automaticamente ou acionada manualmente conforme necessário.
- Retenção de snapshots: os snapshots são versões com carimbo de data e hora de uma tabela do Iceberg. As configurações de retenção de snapshots permitem que os clientes determinem por quanto tempo reter e quantos snapshots devem ser retidos. A configuração de um otimizador de retenção de snapshots pode ajudar a gerenciar a sobrecarga de armazenamento removendo snapshots antigos e desnecessários e seus arquivos subjacentes.
- Exclusão de arquivos órfãos: arquivos órfãos são arquivos que não são mais referidos pelos metadados da tabela do Iceberg. Esses arquivos podem se acumular ao longo do tempo, especialmente após operações como exclusões de tabelas ou trabalhos de ETL com falha. A ativação da exclusão de arquivos órfãos permite identificar e AWS Glue remover periodicamente esses arquivos desnecessários, liberando espaço de armazenamento.

Você pode ativar ou desativar os otimizadores de compactação, retenção de instantâneos e exclusão de arquivos órfãos para tabelas individuais do Iceberg no Catálogo de Dados usando o console ou as operações da API. AWS Glue AWS CLI AWS Glue

Para obter mais informações, consulte Otimizando tabelas Iceberg no Guia do AWS Glue desenvolvedor.

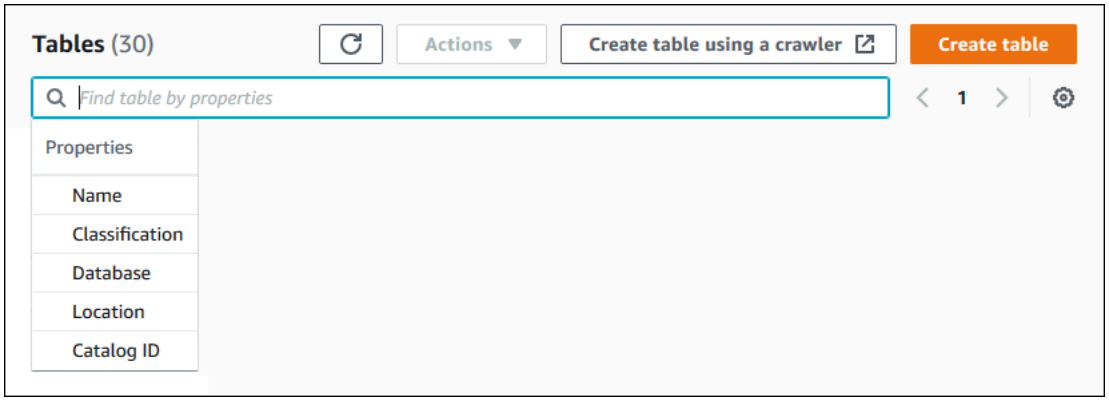
Procurando por tabelas

Você pode usar o AWS Lake Formation console para pesquisar tabelas do Catálogo de Dados por nome, localização, banco de dados contendo e muito mais. Os resultados da pesquisa mostram somente as tabelas nas quais você tem permissões do Lake Formation.

Para procurar tabelas (console)

1. Faça login Console de gerenciamento da AWS e abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.
2. No painel de navegação, selecione Tabelas.
3. Posicione o cursor no campo de pesquisa. O campo tem o texto do espaço reservado Localizar tabela por propriedades.

   O menu Propriedades é exibido, mostrando as várias propriedades da tabela pelas quais pesquisar.



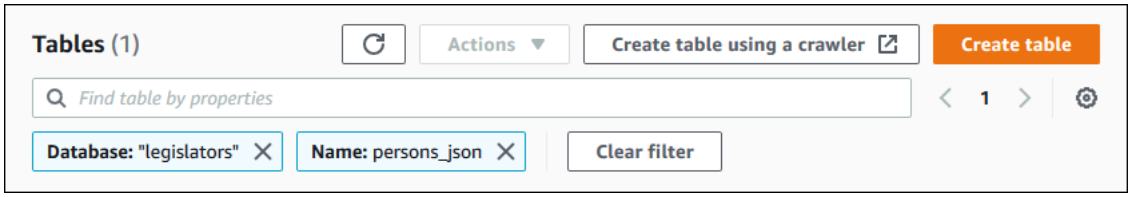
4. Execute um destes procedimentos:

   - Pesquise contendo banco de dados.
     1. Escolha Banco de dados no menu Propriedades e, em seguida, escolha um banco de dados no menu Bancos de dados exibido ou digite o nome do banco de dados e pressione Enter.

As tabelas nas quais você tem permissões no banco de dados são listadas.

2. (Opcional) Para restringir a lista a uma única tabela no banco de dados, posicione o cursor no campo de pesquisa novamente, escolha Nome no menu Propriedades e escolha um nome de tabela no menu Tabelas exibido ou digite um nome de tabela e pressione Enter.

   A tabela única é listada e tanto o nome do banco de dados quanto o nome da tabela aparecem como blocos sob o campo de pesquisa.



   Para ajustar o filtro, feche um dos ladrilhos ou escolha Limpar filtro.

- Pesquise por outras propriedades.

  1. Escolha uma propriedade de pesquisa no menu Propriedades.

     Para pesquisar por ID da AWS conta, escolha ID do catálogo no menu Propriedades, insira uma ID de AWS conta válida (por exemplo, 111122223333) e pressione Enter.

     Para pesquisar por localização, escolha Localização no menu Propriedades e selecione uma localização no menu Localizações que aparece. Todas as tabelas na localização raiz da localização selecionada (por exemplo, Amazon S3) são retornadas.

Pesquisando tabelas usando AWS CLI

- O exemplo a seguir mostra como executar uma pesquisa parcial. O parâmetro `--search-text` permite pesquisar tabelas que contenham o texto especificado em seus metadados. Nesse caso, ele exibe todas as tabelas que têm “cliente” no nome, na descrição ou em outros campos de metadados.

```
aws glue search-tables
      --search-text "customer"
      --region Região da AWS
      --max-results 10
      --sort-criteria "FieldName=Name,Sort=ASC"
```

Compartilhando tabelas e bancos de dados do Catálogo de Dados entre AWS Contas

Você pode compartilhar recursos do Catálogo de Dados (bancos de dados e tabelas) com AWS contas externas concedendo permissões do Lake Formation sobre os recursos às contas externas. Os usuários podem então executar consultas e trabalhos que unem e consultam tabelas em várias contas. Com algumas restrições, quando você compartilha um recurso do catálogo de dados com outra conta, as entidades principais dessa conta podem operar nesse recurso como se o recurso estivesse em seu catálogo de dados.

Você não compartilha recursos com diretores específicos em AWS contas externas — você compartilha os recursos com uma AWS conta ou organização. Ao compartilhar um recurso com uma organização da AWS , você está compartilhando o recurso com todas as contas em todos os níveis dessa organização. O administrador do data lake em cada conta externa deve então conceder permissões sobre os recursos compartilhados às entidades principais da conta.

Para obter mais informações, consulte Cross-account compartilhamento de dados em Lake Formation e Conceder permissões nos recursos do Catálogo de Dados.

Consulte também:

- Acessar e visualizar tabelas e bancos de dados compartilhados do catálogo de dados
- Pré-requisitos

Desenvolvimento AWS Glue Data Catalog visualizações

No AWS Glue Data Catalog, uma exibição é uma tabela virtual na qual o conteúdo é definido por uma consulta SQL que faz referência a uma ou mais tabelas. Você pode criar uma visualização do catálogo de dados que faça referência a até 10 tabelas usando editores SQL para Amazon Athena, Amazon Redshift ou Apache Spark usando o EMR Serverless ou a versão 5.0. AWS Glue As tabelas de referência subjacentes de uma exibição podem pertencer ao mesmo banco de dados ou a bancos de dados diferentes dentro Conta da AWS do mesmo catálogo de dados.

Você pode referenciar AWS Glue tabelas e tabelas padrão em formatos de tabela aberta (OTF), como Apache Hudi, Linux Foundation Delta Lake e Apache Iceberg, com dados subjacentes armazenados em locais do Amazon S3 registrados com. AWS Lake Formation Além disso, você pode criar visualizações de tabelas federadas em unidades de compartilhamento de dados do Amazon Redshift que são compartilhadas com o Lake Formation.

Diferenciar visualizações do Catálogo de Dados de outros tipos de visualização

As visualizações do Catálogo de Dados diferem das visualizações do Apache Hive, do Apache Spark e do Amazon Athena. A visualização do Catálogo de Dados é um recurso nativo do AWS Glue Data Catalog, e é uma visualização criada pelo definidor de vários dialetos. Você pode criar uma visualização do Catálogo de Dados usando um dos serviços de analytics compatíveis, como o Athena ou o Amazon Redshift Spectrum, e acessar a mesma visualização usando outros serviços de analytics compatíveis. Entretanto, as visualizações do Apache Hive, do Apache Spark e do Athena são criadas de forma independente em cada serviço de analytics, como o Athena e o Amazon Redshift, e são visíveis e acessíveis somente dentro desse serviço.

O que é uma visualização de programador?

Uma visualização de programador é uma visualização SQL que opera com base nas permissões da entidade principal que a criou. O perfil do programador tem as permissões necessárias para acessar as tabelas referidas e executa a instrução SQL que programa a visualização. O definidor cria a visualização e a compartilha com outros usuários por meio AWS Lake Formation do controle de acesso refinado.

Quando um usuário consulta a visualização do programador, o mecanismo de consulta usa as permissões do perfil do programador para acessar as tabelas de referência subjacentes. Essa abordagem permite que os usuários interajam com a visualização sem precisar de acesso direto às tabelas de origem, aumentando a segurança e simplificando o gerenciamento do acesso aos dados.

Para configurar uma visualização definidora, a função definidora do IAM pode estar na mesma AWS conta das tabelas base ou em uma conta diferente usando funções de definidor entre contas. Para obter mais informações sobre as permissões necessárias para o perfil do programador, consulte Pré-requisitos para criar visualizações.

Um framework para visualizações de vários dialetos

O Catálogo de Dados aceita a criação de visualizações usando vários dialetos da linguagem de consulta estruturada (SQL). SQL é uma linguagem usada para armazenar e processar informações em um banco de dados relacional e cada mecanismo AWS analítico usa sua própria variação de SQL ou dialeto SQL.

Você cria uma visualização do Catálogo de Dados em um dialeto SQL usando um dos mecanismos de consulta de analytics compatíveis. Posteriormente, você pode atualizar a visualização usando a instrução `ALTER VIEW` em um dialeto SQL diferente em qualquer outro mecanismo de analytics

compatível. No entanto, cada dialeto deve fazer referência ao mesmo conjunto de tabelas, colunas e tipos de dados.

Você pode acessar os vários dialetos disponíveis para a visualização usando a `GetTable` API AWS CLI e AWS o console. Assim, a visualização do Catálogo de Dados fica visível e disponível para consultas em diferentes mecanismos de analytics compatíveis.

Ao definir um esquema de visualização comum e um objeto de metadados que você pode consultar em vários mecanismos, as visualizações do catálogo de dados permitem usar visualizações uniformes em todo o data lake.

Para obter mais detalhes sobre como o esquema é resolvido para cada dialeto, consulte o link para a referência da API. Para obter mais detalhes sobre as regras de correspondência para diferentes tipos, consulte o link para a seção relevante no documento da API.

Integrar a permissões do Lake Formation

Você pode usar AWS Lake Formation para centralizar o gerenciamento de permissões nas AWS Glue Data Catalog visualizações dos usuários. Você pode conceder permissões refinadas nas visualizações do Catálogo de Dados usando o método de recurso nomeado ou LF-Tags compartilhá-las entre AWS organizações e Contas da AWS unidades organizacionais. Você também pode compartilhar e acessar as visualizações do Catálogo de Dados Regiões da AWS usando links de recursos. Isso permite que os usuários forneçam acesso aos dados sem duplicar a fonte de dados ou compartilhar as tabelas subjacentes.

A declaração `CREATE VIEW` DDL de uma visualização do catálogo de dados pode referenciar AWS Glue as tabelas e tabelas padrão em formatos de tabela aberta (OTF), como Hudi, Delta Lake e Iceberg, com dados subjacentes armazenados em locais do Amazon S3 registrados no Lake Formation, bem como as tabelas federadas do compartilhamento de dados do Amazon Redshift que são compartilhadas com o Lake Formation. As tabelas podem ter qualquer formato de arquivo, desde que o mecanismo usado para consultar a visualização seja compatível com esse formato. Você também pode fazer referência a funções integradas do mecanismo no qual elas são executadas, mas outros recursos específicos do mecanismo podem não ser permitidos. Para obter mais detalhes, consulte Considerações e limitações das visualizações do catálogo de dados

Casos de uso

Os casos de uso importantes das visualizações do Catálogo de Dados são apresentados abaixo:

- Criar e gerenciar permissões em um único esquema de visualização. Isso ajuda a evitar o risco de permissões inconsistentes em visualizações duplicadas criadas em vários mecanismos.
- Conceda permissões aos usuários em uma visualização que faz referência a várias tabelas sem conceder permissões diretamente nas tabelas de referência subjacentes.
- Obtenha a filtragem em nível de linha em tabelas usando LF-Tags (onde somente em LF-Tags cascata até o nível da coluna) aplicando LF-Tags em visualizações e concedendo permissões LF-Tags baseadas aos usuários.

Compatível AWS serviços de análise para visualizações

Os seguintes serviços de AWS análise oferecem suporte à criação de visualizações do Catálogo de Dados:

- banco de dados de origem
- Amazon Athena versão 3
- Apache Spark no EMR Sem Servidor
- Apache Spark na versão 5.0 AWS Glue

Recursos adicionais do

É possível saber mais sobre o Catálogo de Dados neste guia do usuário, bem como nos seguintes recursos:

O vídeo a seguir demonstra como criar visualizações e consultá-las no Athena e no Amazon Redshift.

Tópicos

- Pré-requisitos para criar visualizações
- Criar visualizações do Catálogo de Dados usando instruções DDL
- Criando visualizações do Catálogo de Dados usando AWS Glue APIs
- Conceder permissões nas visualizações do catálogo de dados
- Visões materializadas

Pré-requisitos para criar visualizações

- Para criar visualizações no catálogo de dados, é necessário registrar os locais de dados subjacentes do Amazon S3 das tabelas de referência no Lake Formation. Para obter detalhes sobre o registro de dados no Lake Formation, consulte Adicionar uma localização do Amazon S3 ao seu data lake.
- Somente perfis do IAM podem criar visualizações do Catálogo de Dados. Outras identidades do IAM não podem criar visualizações do catálogo de dados.
- O perfil do IAM que define a visualização deve ter as seguintes permissões:
  - Permissão `SELECT` completa do Lake Formation com a opção `Grantable` em todas as tabelas de referência, com todas as colunas incluídas.
  - A permissão `CREATE_TABLE` do Lake Formation no banco de dados de destino em que as visualizações estão sendo criadas.
  - Uma política de confiança para que a Lake Formation e AWS Glue os serviços assumam a função.

    JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "DataCatalogViewDefinerAssumeRole1",
            "Effect": "Allow",
            "Principal": {
               "Service": [
                    "glue.amazonaws.com",
                    "lakeformation.amazonaws.com"
                ]
            },
            "Action": "sts:AssumeRole"
        }
    ]
}
```

  - O objetivo: PassRole permissão para AWS Glue e Lake Formation.

JSON

```
{
    "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 e permissões do Lake Formation.

  JSON

```
{
    "Version":"2012-10-17",
                 "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "Glue:GetDatabase",
                "Glue:GetDatabases",
                "Glue:CreateTable",
                "Glue:GetTable",
                "Glue:GetTables",
                "Glue:BatchGetPartition",
                "Glue:GetPartitions",
                "Glue:GetPartition",
                "Glue:GetTableVersion",
```

```
                "Glue:GetTableVersions",
    "Glue:PassConnection",
                "lakeFormation:GetDataAccess"
            ],
            "Resource": "*"
        }
    ]
}
```

- Não é possível criar visualizações em um banco de dados que tenha a permissão `Super` ou `ALL` concedida ao grupo `IAMAllowedPrincipals`. Você pode revogar a permissão `Super` para o grupo `IAMAllowedPrincipals` em um banco de dados, consultar Etapa 4: mude seus armazenamentos de dados para o modelo de permissões do Lake Formation ou criar um banco de dados com a caixa Usar somente o controle de acesso do IAM para novas tabelas neste banco de dados desmarcada em Permissões padrão para tabelas recém-criadas.

Criar visualizações do Catálogo de Dados usando instruções DDL

Você pode criar AWS Glue Data Catalog visualizações usando editores SQL para Athena, Amazon Redshift e usando as APIs/. AWS Glue AWS CLI

Para criar uma visualização do Catálogo de Dados usando editores SQL, escolha Athena ou Redshift Spectrum e crie a visualização usando uma instrução `CREATE VIEW` da linguagem de definição de dados (DDL). Depois de criar uma visualização no dialeto do primeiro mecanismo, você pode usar uma instrução DDL `ALTER VIEW` do segundo mecanismo para adicionar os outros dialetos.

Ao definir as visualizações, é importante considerar o seguinte:

- Definir visualizações de vários dialetos: quando você define uma visualização com vários dialetos, os esquemas dos diferentes dialetos devem corresponder. Cada dialeto SQL terá uma especificação de sintaxe ligeiramente diferente. A sintaxe da consulta que define a visualização do Catálogo de Dados deve ser resolvida exatamente na mesma lista de colunas, incluindo tipos e nomes, em todos os dialetos. Essas informações são armazenadas no `StorageDescriptor` da visualização. Os dialetos também devem fazer referência aos mesmos objetos subjacentes da tabela do Catálogo de Dados.

  Para adicionar outro dialeto a uma visualização usando DDL, você pode usar a instrução `ALTER VIEW`. Se uma instrução `ALTER VIEW` tentar atualizar a definição da visualização, como modificar o descritor de armazenamento ou as tabelas subjacentes da visualização, a instrução vai gerar a

mensagem de erro: “Input and existing storage descriptor mismatch”. Você pode usar operações de conversão de SQL para garantir que os tipos de coluna de visualização correspondam.

- Atualizar uma visualização: para atualizar a visualização, você pode usar a API `UpdateTable`. Se você atualizar a visualização sem compatibilizar os descritores de armazenamento ou as tabelas de referência, poderá fornecer o sinalizador `FORCE` (consulte a documentação do SQL do mecanismo para obter a sintaxe). Depois de uma atualização forçada, a visualização assumirá a tabela `StorageDescriptor` forçada e a de referência. Qualquer DDL `ALTER VIEW` adicional deve corresponder aos valores modificados. Uma visualização que foi atualizada para ter dialetos incompatíveis terá o status “Obsoleto”. O status da visualização é visível no console do Lake Formation e usa a operação `GetTable`.
- Fazer referência a um tipo de coluna varchar como uma string: não é possível converter um tipo de coluna varchar do Redshift Spectrum em uma string. Se uma visualização for criada no Redshift Spectrum com um tipo de coluna varchar e um dialeto subsequente tentar fazer referência a esse campo como uma string, o Catálogo de Dados a tratará como string sem a necessidade do sinalizador `FORCE`.
- Tratamento de campos de tipos complexos: o Amazon Redshift trata todos os tipos complexos como tipos SUPER, enquanto o Athena especifica o tipo complexo. Se uma visualização tiver um campo de tipo `SUPER` e outro mecanismo fizer referência a essa coluna como um tipo complexo específico, como struct (`<street_address:struct<street_number:int, street_name:string, street_type:string>>`), o Catálogo de Dados presumirá que o campo seja do tipo complexo específico e o usará no descritor de armazenamento, sem exigir o sinalizador `Force`.

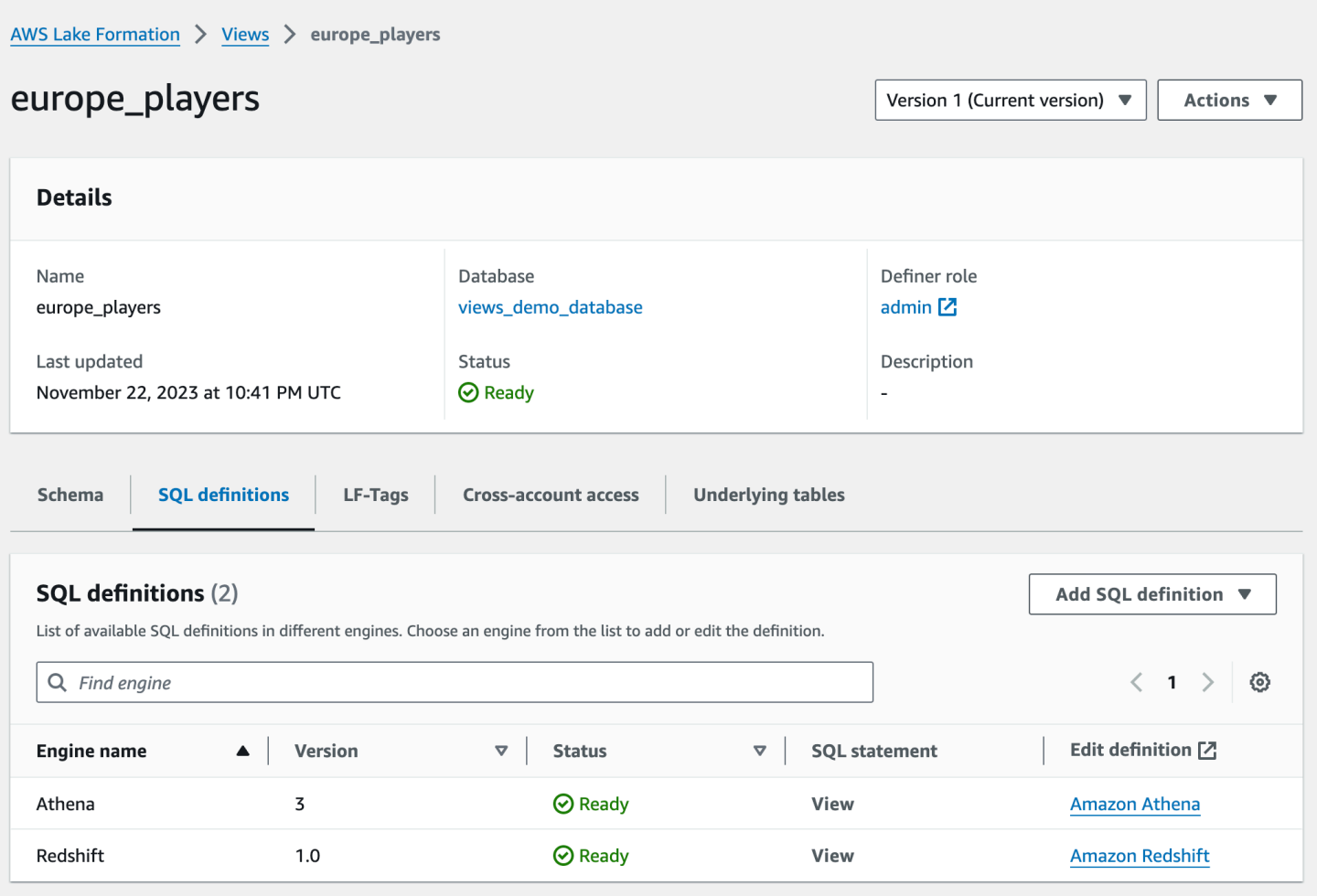
Para obter mais informações sobre a sintaxe para criar e gerenciar visualizações do catálogo de dados, consulte:

- Uso das visualizações do AWS Glue Data Catalog no Guia do usuário do Amazon Athena.
- Glue Data Catalog view query syntax no Guia do usuário do Amazon Athena.
- Creating views in the AWS Glue Data Catalog no Guia do desenvolvedor do banco de dados do Amazon Redshift.

  Para obter mais informações sobre os comandos SQL relacionados a exibições no Data Catalog, consulte CREATE EXTERNAL VIEW, ALTER EXTERNAL VIEW e DROP EXTERNAL VIEW.

Após a criação de uma visualização do Catálogo de Dados, os respectivos detalhes ficam disponíveis no console do Lake Formation.

1. Selecione Visualizações no catálogo de dados no console do Lake Formation.
2. Uma lista das visualizações disponíveis é exibida na página de visualizações.
3. Selecione uma visualização na lista e a página de detalhes mostrará os atributos da visualização.



Schema

Escolha uma `Column` linha e selecione Editar LF-Tags para atualizar os valores da tag ou atribuir novos LF-Tags.

Definições de SQL

É possível ver uma lista das definições de SQL disponíveis. Selecione Adicionar definição de SQL e escolha um mecanismo de consulta para adicionar uma definição de SQL. Selecione

um mecanismo de consulta (Athena ou Amazon Redshift) na coluna `Edit definition` para atualizar as definições de SQL.

LF-Tags

Escolha Editar LF-Tags para editar valores para uma tag ou atribuir novas tags. Você pode usar LF-Tags para conceder permissões nas visualizações.

Cross-account acesso

Você pode ver uma lista de Contas da AWS organizações e unidades organizacionais (OUs) com as quais você compartilhou a visualização do Catálogo de Dados.

Tabelas subjacentes

As tabelas subjacentes referenciadas na definição de SQL usada para criar a exibição são mostradas nessa guia.

Criando visualizações do Catálogo de Dados usando AWS Glue APIs

Você pode usar AWS Glue CreateTablee UpdateTableAPIs para criar e atualizar visualizações no Catálogo de Dados. As operações `CreateTable` e `UpdateTable` têm uma nova estrutura de `TableInput` para `ViewDefinition`, enquanto as operações `SearchTables`, `GetTable`, `GetTables`, `GetTableVersion` e `GetTableVersions` fornecem a `ViewDefinition` na respectiva sintaxe de saída para as visualizações. Além disso, há um novo campo de `Status` na saída da API `GetTable`.

Duas novas AWS Glue conexões estão disponíveis para validar o dialeto SQL para cada mecanismo de consulta compatível Amazon Athena e para o Amazon Redshift.

As APIs `CreateTable` e `UpdateTable` são assíncronas quando usadas com visualizações. Quando essas APIs são chamadas com vários dialetos SQL, a chamada é validada com cada mecanismo para determinar se o dialeto pode ser executado nesse mecanismo e se o esquema resultante da visualização de cada dialeto corresponde. O AWS Glue serviço usa essas conexões para fazer chamadas internas para os mecanismos analíticos. Essas chamadas simulam o que o mecanismo faz para validar se uma instrução DDL SQL `CREATE VIEW` ou `ALTER VIEW` foi executada no mecanismo.

Se o SQL fornecido for válido e os esquemas corresponderem entre os dialetos de visualização, a API do AWS Glue confirmará atomicamente o resultado. A atomicidade permite que visualizações com vários dialetos sejam criadas ou alteradas sem nenhum tempo de inatividade.

Tópicos

- Criando AWS Glue conexões para validar o status
- Validar o status de geração de visualizações
- Estados e operações assíncronos
- Cenários de falha na criação de visualizações durante operações assíncronas

Criando AWS Glue conexões para validar o status

Para criar ou atualizar uma AWS Glue Data Catalog visualização usando as `UpdateTable` operações `CreateTable` ou, você deve criar um novo tipo de AWS Glue conexão para validação e fornecê-lo ao mecanismo de análise compatível. Essas conexões são necessárias para usar as visualizações do Catálogo de Dados com o Athena ou o Amazon Redshift. Você só pode criar essas conexões usando AWS SDKs ou AWS Glue APIs. AWS CLI Você não pode usar o Console de gerenciamento da AWS para criar a AWS Glue conexão.

> **Note**
>
> Se o perfil do programador de visualizações e o perfil que chama `CreateTable` ou `UpdateTable` forem diferentes, ambos precisarão da permissão `glue:PassConnection` na declaração de política do IAM.

Para obter mais informações, consulte a documentação de criação de conexão AWS CLI .

AWS CLI comando para criar uma conexão

Veja a seguir um AWS CLI comando para criar uma conexão:

```
aws glue create-connection --region us-east-1
--endpoint-url https://glue.us-east-1.amazonaws.com
--cli-input-json file:///root/path/to/create-connection.json
```

AWS CLI entrada JSON

Para Amazon Redshift:

```
{
    "CatalogId": "123456789012",
    "ConnectionInput": {
```

```
        "ConnectionType": "VIEW_VALIDATION_REDSHIFT",
        "Name": "views-preview-cluster-connection-2",
        "Description": "My first Amazon Redshift validation connection",
        "ConnectionProperties": {
            "DATABASE": "dev",
            "CLUSTER_IDENTIFIER": "glue-data-catalog-views-preview-cluster"
        }
    }
}
```

Para Amazon Athena:

```
{
    "CatalogId": "123456789012",
    "ConnectionInput": {
        "ConnectionType": "VIEW_VALIDATION_ATHENA",
        "Name": "views-preview-cluster-connection-3",
        "Description": "My first Amazon Athena validation connection",
        "ConnectionProperties": {
            "WORKGROUP_NAME": "workgroup-name"
        }
    }
}
```

Validar o status de geração de visualizações

Quando você executa as operações `CreateTable` ou `UpdateTable`, o campo `Status` da saída da API `GetTable` mostra os detalhes do status de criação da visualização. Para `create` solicitações em que a tabela ainda não existe, AWS Glue cria uma tabela vazia durante o processo assíncrono. Ao chamar `GetTable`, você pode passar um sinalizador booliano `IncludeStatusDetails` opcional, que mostra informações de diagnóstico sobre a solicitação. No caso de uma falha, esse sinalizador mostra uma mensagem de erro com o status individual de cada dialeto.

Erros durante as operações de criação, leitura, atualização e exclusão de visualizações (CRUD) podem ocorrer durante o processamento no serviço AWS Glue/Lake Formation ou durante a validação do SQL de visualização no Amazon Redshift ou no Athena. Quando ocorre um erro durante a validação em um mecanismo, o serviço AWS Glue apresenta a mensagem de erro que o mecanismo exibe.

Campos de status

Estes são os campos de status:

- Status (um status genérico, que se aplica a diferentes tipos de trabalho):
  - QUEUED
  - IN_PROGRESS
  - SUCCESS
  - FAILED
- Action: indica qual ação foi chamada na tabela. No momento, somente as operações `CREATE` ou `UPDATE` estão disponíveis.

  Distinguir entre as operações `UPDATE` e `CREATE` é importante ao trabalhar com visualizações. O tipo de operação determina como você deve prosseguir com a consulta das tabelas.

  Uma operação `UPDATE` significa que a tabela já existe no Catálogo de Dados. Nesse caso, você pode continuar consultando a tabela criada anteriormente sem problemas. Entretanto, uma operação `CREATE` indica que a tabela nunca foi criada com sucesso. Se uma tabela estiver marcada como `CREATE`, a tentativa de consultá-la falhará porque ela ainda não existe no sistema. Portanto, é essencial identificar o tipo de operação (UPDATE ou CREATE) antes de tentar consultar uma tabela.
- RequestedBy — O ARN do usuário que solicitou a alteração assíncrona.
- UpdatedBy — O ARN do usuário que alterou manualmente pela última vez o processo de alteração assíncrona, como solicitar um cancelamento ou modificação.
- Error: este campo só aparece quando o estado é FAILED. Essa é uma mensagem de exceção no nível principal. Pode haver erros diferentes para cada dialeto.
  - ErrorCode — O tipo de exceção.
  - ErrorMessage — uma breve descrição da exceção.
- RequestTime — uma string de data formatada em ISO 8601 indicando a hora em que a alteração foi iniciada.
- UpdateTime — uma string de data formatada em ISO 8601 indicando a hora em que o estado foi atualizado pela última vez.

Estados e operações assíncronos

Quando você executa uma solicitação `glue:CreateTable`, a criação assíncrona da visualização do Catálogo de Dados inicia-se. Nas seções a seguir, este documento descreve `Status` a AWS Glue visão que está disponível em uma `glue:GetTable` resposta. Por motivo de brevidade, esta seção omite a resposta completa.

```
{
    "Table": {
        ...
        "Status": {
            ...
            "Action": "CREATE",
            "State": "QUEUED",
        }
    }
}
```

Ambos os atributos acima representam informações importantes de diagnóstico que indicam o estado da operação assíncrona, bem como as ações que podem ser executadas nessa visualização. Abaixo estão os valores possíveis que esses atributos podem assumir.

1. `Status.Action`
   a. CREATE
   b. UPDATE
2. `Status.State`
   a. QUEUED
   b. IN_PROGRESS
   c. SUCCESS
   d. FAILED

Também é importante observar que algumas atualizações em uma visualização do Catálogo de Dados não exigem uma operação assíncrona. Por exemplo, para atualizar o atributo `Description` da tabela. Como isso não requer nenhuma operação assíncrona, os metadados da tabela resultante não terão nenhum `Status`, e o atributo será `NULL`.

```
{
    "Table": {
        ...,
        "Description": "I changed this attribute!"
    }
}
```

A seguir, este tópico explora como as informações de status acima podem afetar as operações que podem ser executadas em uma AWS Glue exibição.

cola: CreateTable

Não há alterações nessa API em comparação com a forma como `glue:CreateTable` funciona para qualquer tabela do Glue. É possível chamar `CreateTable` para qualquer nome de tabela que ainda não exista.

cola: UpdateTable

Essa operação não pode ser executada em uma AWS Glue exibição que tenha as seguintes informações de status:

1. Action == CREATE e State == QUEUED
2. Action == CREATE e State == IN_PROGRESS
3. Action == CREATE e State == FAILED
4. Action == UPDATE e State == QUEUED
5. Action == UPDATE e State == IN_PROGRESS

Resumindo, você pode atualizar uma visualização do Catálogo de Dados somente quando ela atender aos requisitos a seguir.

1. Foi criada com sucesso pela primeira vez.
   a. Action == CREATE e State == SUCCESS
2. Atingiu um estado final após uma operação de atualização assíncrona.
   a. Action == UPDATE e State == SUCCESS
   b. Action == UPDATE e State == FAILED
3. Tem um atributo de estado `NULL` em decorrência de uma atualização síncrona.

cola: DeleteTable

Não há alterações nessa operação quando comparada à forma como `glue:DeleteTable` funciona em qualquer AWS Glue tabela. Você pode excluir uma visualização do Catálogo de Dados, independentemente do respectivo estado.

cola: GetTable

Não há alterações nessa operação quando comparada à forma como `glue:GetTable` funciona em qualquer AWS Glue tabela. No entanto, não é possível consultar uma visualização do Catálogo de Dados nos mecanismos analíticos enquanto ela não for criada com sucesso pela primeira vez. `Action == CREATE and State == SUCCESS`. Depois de criar uma visualização do Catálogo de Dados com sucesso pela primeira vez, você pode consultar a visualização, independentemente do respectivo status.

> **Note**
>
> Todas as informações nesta seção se aplicam a todas as APIs de leitura de tabelas, tais como `GetTable`, `GetTables` e `SearchTables`.

Cenários de falha na criação de visualizações durante operações assíncronas

Os exemplos a seguir são representativos dos tipos de erro que podem resultar das chamadas de API `CreateTable` ou `UpdateTable` da visualização. Eles não são completos, pois a superfície de erro das falhas de consultas SQL é muito grande.

Cenário 1: falha na consulta do Amazon Redshift

A consulta fornecida para o Amazon Redshift inclui um nome de tabela com erro de ortografia que não pode ser encontrado no Catálogo de Dados durante a validação. O erro resultante é mostrado no campo `Status` na resposta `GetTable` da visualização.

Solicitação `GetTable`:

```
{
    "CatalogId": "123456789012",
    "DatabaseName": "async-view-test-db",
    "TableInput": {
        "Name": "view-athena-redshift-72",
        "Description": "This is an atomic operation",
        "StorageDescriptor": {
            "Columns": [
                { "Name": "col1", "Type": "int" },
                { "Name": "col2", "Type": "string" },
                { "Name": "col3", "Type": "double" }
            ]
        },
        "ViewDefinition": {
```

```
            "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
            "SubObjects": [ "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1" ],
            "Representations": [
                {
                    "Dialect": "ATHENA",
                    "DialectVersion": "3",
                    "ViewOriginalText": "SELECT * FROM \"gdc-view-playground-db\".
\"table_1\"",
                    "ValidationConnection": "athena-connection"
                },
                {
                    "Dialect": "REDSHIFT",
                    "DialectVersion": "1.0",
                    "ViewOriginalText": "SELECT * FROM \"gdc-view-playground-external-
schema\".\"table__1\";",
                    "ValidationConnection": "redshift-connection"
                }
            ]
        }
    }
}
```

Resposta `GetTable`:

```
IncludeStatusDetails = FALSE
{
    "Table": {
        "Name": "view-athena-redshift-72",
        "DatabaseName": "async-view-test-db",
        "Description": "",
        "CreateTime": "2024-07-11T11:39:19-07:00",
        "UpdateTime": "2024-07-11T11:39:19-07:00",
        "Retention": 0,
        "ViewOriginalText": "",
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
```

```
        "IsMultiDialectView": false,
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:39:19-07:00",
            "UpdateTime": "2024-07-11T11:40:06-07:00",
            "Action": "CREATE",
            "State": "FAILED"
        }
    }
}

IncludeStatusDetails = TRUE
{
    "Table": {
        "Name": "view-athena-redshift-72",
        "DatabaseName": "async-view-test-db",
        "Description": "",
        "CreateTime": "2024-07-11T11:39:19-07:00",
        "UpdateTime": "2024-07-11T11:39:19-07:00",
        "Retention": 0,
        "ViewOriginalText": "",
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "IsMultiDialectView": false,
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:39:19-07:00",
            "UpdateTime": "2024-07-11T11:40:06-07:00",
            "Action": "CREATE",
            "State": "FAILED",
            "Error": {
                "ErrorCode": "QueryExecutionException",
                "ErrorMessage": "Error received during view SQL validation
 using a connection: [Connection Name: redshift-connection | Query Execution
 Id: ddb711d3-2415-4aa9-b251-6a76ab4f41b1 | Timestamp: Thu Jul 11 18:39:37 UTC
```

```
 2024]: Redshift returned error for the statement: ERROR: AwsClientException:
 EntityNotFoundException from glue - Entity Not Found"
            },
            "Details": {
                "RequestedChange": {
                    "Name": "view-athena-redshift-72",
                    "DatabaseName": "async-view-test-db",
                    "Description": "This is an atomic operation",
                    "Retention": 0,
                    "StorageDescriptor": {
                        "Columns": [
                            {
                                "Name": "col1",
                                "Type": "int"
                            },
                            {
                                "Name": "col2",
                                "Type": "string"
                            },
                            {
                                "Name": "col3",
                                "Type": "double"
                            }
                        ],
                        "Compressed": false,
                        "NumberOfBuckets": 0,
                        "SortColumns": [],
                        "StoredAsSubDirectories": false
                    },
                    "TableType": "VIRTUAL_VIEW",
                    "IsRegisteredWithLakeFormation": false,
                    "CatalogId": "123456789012",
                    "IsRowFilteringEnabled": false,
                    "VersionId": "-1",
                    "DatabaseId": "<databaseID>",
                    "ViewDefinition": {
                        "IsProtected": true,
                        "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
                        "SubObjects": [
                            "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1"
                        ],
                        "Representations": [
                            {
```

```
                                "Dialect": "ATHENA",
                                "DialectVersion": "3",
                                "ViewOriginalText": "SELECT * FROM \"gdc-view-
playground-db\".\"table_1\"",
                                "IsStale": false
                            },
                            {
                                "Dialect": "REDSHIFT",
                                "DialectVersion": "1.0",
                                "ViewOriginalText": "SELECT * FROM \"gdc-view-
playground-external-schema\".\"table__1\";",
                                "IsStale": false
                            }
                        ]
                    },
                    "IsMultiDialectView": true
                },
                "ViewValidations": [
                    {
                        "Dialect": "ATHENA",
                        "DialectVersion": "3",
                        "ViewValidationText": "SELECT * FROM \"gdc-view-playground-db
\".\"table_1\"",
                        "UpdateTime": "2024-07-11T11:40:06-07:00",
                        "State": "SUCCESS"
                    },
                    {
                        "Dialect": "REDSHIFT",
                        "DialectVersion": "1.0",
                        "ViewValidationText": "SELECT * FROM \"gdc-view-playground-
external-schema\".\"table__1\";",
                        "UpdateTime": "2024-07-11T11:39:37-07:00",
                        "State": "FAILED",
                        "Error": {
                            "ErrorCode": "QueryExecutionException",
                            "ErrorMessage": "Error received during view SQL validation
 using a connection: [Connection Name: redshift-connection | Query Execution Id:
 ddb711d3-2415-4aa9-b251-6a76ab4f41b1 | Timestamp: Thu
 Jul 11 18:39:37 UTC 2024]: Redshift returned error for the statement: ERROR:
 AwsClientException: EntityNotFoundException from glue - Entity Not Found"
                        }
                    }
                ]
            }
```

```
        }
    }
}
```

Cenário 2: conexão inválida do Amazon Redshift

A conexão do Amazon Redshift no exemplo a seguir está malformada porque se refere a um banco de dados do Amazon Redshift que não existe no endpoint fornecido. cluster/serverless O Amazon Redshift não consegue validar a visualização e o campo `Status` na resposta `GetTable` mostra o erro `"State": "FAILED"` do Amazon Redshift).

Solicitação `GetTable`:

```
{
    "CatalogId": "123456789012",
    "DatabaseName": "async-view-test-db",
    "TableInput": {
        "Name": "view-athena-redshift-73",
        "Description": "This is an atomic operation",
        "StorageDescriptor": {
            "Columns": [
                { "Name": "col1", "Type": "int" },
                { "Name": "col2", "Type": "string" },
                { "Name": "col3", "Type": "double" }
            ]
        },
        "ViewDefinition": {
            "Definer": "arn:aws:iam::123456789012:role/GDCViewDdefiner",
            "SubObjects": [ "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1" ],
            "Representations": [
                {
                    "Dialect": "ATHENA",
                    "DialectVersion": "3",
                    "ViewOriginalText": "SELECT * FROM \"gdc-view-playground-db\".
\"table_1\"",
                    "ValidationConnection": "athena-connection"
                },
                {
                    "Dialect": "REDSHIFT",
                    "DialectVersion": "1.0",
                    "ViewOriginalText": "SELECT * FROM \"gdc-view-playground-external-
schema\".\"table_1\";",
```

```
                    "ValidationConnection": "redshift-connection-malformed"
                }
            ]
        }
    }
}
```

Resposta `GetTable`:

```
IncludeStatusDetails = FALSE
{
    "Table": {
        "Name": "view-athena-redshift-73",
        "DatabaseName": "async-view-test-db",
        "Description": "",
        "CreateTime": "2024-07-11T11:43:27-07:00",
        "UpdateTime": "2024-07-11T11:43:27-07:00",
        "Retention": 0,
        "ViewOriginalText": "",
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "IsMultiDialectView": false,
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:43:27-07:00",
            "UpdateTime": "2024-07-11T11:43:40-07:00",
            "Action": "CREATE",
            "State": "FAILED"
        }
    }
}

IncludeStatusDetails = TRUE
{
    "Table": {
        "Name": "view-athena-redshift-73",
```

```
        "DatabaseName": "async-view-test-db",
        "Description": "",
        "CreateTime": "2024-07-11T11:43:27-07:00",
        "UpdateTime": "2024-07-11T11:43:27-07:00",
        "Retention": 0,
        "ViewOriginalText": "",
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "IsMultiDialectView": false,
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:43:27-07:00",
            "UpdateTime": "2024-07-11T11:43:40-07:00",
            "Action": "CREATE",
            "State": "FAILED",
            "Error": {
                "ErrorCode": "QueryExecutionException",
                "ErrorMessage": "Error received during view SQL validation using a
connection: [Connection Name: redshift-connection-malformed | Query Execution Id:
69bfafd4-3d51-4cb0-9320-7ce5404b1809 | Timestamp: Thu Jul 11 18:43:38 UTC 2024]:
Redshift returned error for the statement: FATAL: database \"devooo\" does not exist"
            },
            "Details": {
                "RequestedChange": {
                    "Name": "view-athena-redshift-73",
                    "DatabaseName": "async-view-test-db",
                    "Description": "This is an atomic operation",
                    "Retention": 0,
                    "StorageDescriptor": {
                        "Columns": [
                            {
                                "Name": "col1",
                                "Type": "int"
                            },
                            {
                                "Name": "col2",
                                "Type": "string"
```

```
                            },
                            {
                                "Name": "col3",
                                "Type": "double"
                            }
                        ],
                        "Compressed": false,
                        "NumberOfBuckets": 0,
                        "SortColumns": [],
                        "StoredAsSubDirectories": false
                    },
                    "TableType": "VIRTUAL_VIEW",
                    "IsRegisteredWithLakeFormation": false,
                    "CatalogId": "123456789012",
                    "IsRowFilteringEnabled": false,
                    "VersionId": "-1",
                    "DatabaseId": "<databaseID>",
                    "ViewDefinition": {
                        "IsProtected": true,
                        "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
                        "SubObjects": [
                            "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1"
                        ],
                        "Representations": [
                            {
                                "Dialect": "ATHENA",
                                "DialectVersion": "3",
                                "ViewOriginalText": "SELECT * FROM \"gdc-view-
playground-db\".\"table_1\"",
                                "IsStale": false
                            },
                            {
                                "Dialect": "REDSHIFT",
                                "DialectVersion": "1.0",
                                "ViewOriginalText": "SELECT * FROM \"gdc-view-
playground-external-schema\".\"table_1\";",
                                "IsStale": false
                            }
                        ]
                    },
                    "IsMultiDialectView": true
                },
                "ViewValidations": [
```

```
                    {
                        "Dialect": "ATHENA",
                        "DialectVersion": "3",
                        "ViewValidationText": "SELECT * FROM \"gdc-view-playground-db
\".\"table_1\"",
                        "UpdateTime": "2024-07-11T11:43:40-07:00",
                        "State": "SUCCESS"
                    },
                    {
                        "Dialect": "REDSHIFT",
                        "DialectVersion": "1.0",
                        "ViewValidationText": "SELECT * FROM \"gdc-view-playground-
external-schema\".\"table_1\";",
                        "UpdateTime": "2024-07-11T11:43:38-07:00",
                        "State": "FAILED",
                        "Error": {
                            "ErrorCode": "QueryExecutionException",
                            "ErrorMessage": "Error received during view SQL validation
 using a connection: [Connection Name: redshift-connection-malformed | Query Execution
 Id: 69bfafd4-3d51-4cb0-9320-7ce5404b1809 | Time
stamp: Thu Jul 11 18:43:38 UTC 2024]: Redshift returned error for the statement: FATAL:
 database \"devooo\" does not exist"
                        }
                    }
                ]
            }
        }
    }
}
```

Cenário 3: falha na consulta do Athena

Aqui, o SQL para o Athena é inválido porque a consulta digita incorretamente o nome do banco de dados. A validação da consulta do Athena detecta isso e o erro resultante aparece por meio do objeto `Status` em uma chamada `GetTable`.

Solicitação `GetTable`:

```
{
    "CatalogId": "123456789012",
    "DatabaseName": "async-view-test-db",
```

```
    "TableInput": {
        "Name": "view-athena-redshift-70",
        "Description": "This is an atomic operation",
        "StorageDescriptor": {
            "Columns": [
                { "Name": "col1", "Type": "int" },
                { "Name": "col2", "Type": "string" },
                { "Name": "col3", "Type": "double" }
            ]
        },
        "ViewDefinition": {
            "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
            "SubObjects": [ "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1" ],
            "Representations": [
                {
                    "Dialect": "ATHENA",
                    "DialectVersion": "3",
                    "ViewOriginalText": "SELECT * FROM \"gdc--view-playground-db\".
\"table_1\"",
                    "ValidationConnection": "athena-connection"
                },
                {
                    "Dialect": "REDSHIFT",
                    "DialectVersion": "1.0",
                    "ViewOriginalText": "SELECT * FROM \"gdc-view-playground-external-
schema\".\"table_1\";",
                    "ValidationConnection": "redshift-connection"
                }
            ]
        }
    }
}
```

Resposta `GetTable`:

```
IncludeStatusDetails = FALSE
{
    "Table": {
        "Name": "view-athena-redshift-70",
        "DatabaseName": "async-view-test-db",
        "Description": "",
        "CreateTime": "2024-07-11T11:09:53-07:00",
```

```
        "UpdateTime": "2024-07-11T11:09:53-07:00",
        "Retention": 0,
        "ViewOriginalText": "",
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "IsMultiDialectView": false,
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:09:54-07:00",
            "UpdateTime": "2024-07-11T11:10:41-07:00",
            "Action": "CREATE",
            "State": "FAILED",
        }
    }
}

IncludeStatusDetails = TRUE
{
    "Table": {
        "Name": "view-athena-redshift-70",
        "DatabaseName": "async-view-test-db",
        "Description": "",
        "CreateTime": "2024-07-11T11:09:53-07:00",
        "UpdateTime": "2024-07-11T11:09:53-07:00",
        "Retention": 0,
        "ViewOriginalText": "",
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "IsMultiDialectView": false,
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
```

```
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:09:54-07:00",
            "UpdateTime": "2024-07-11T11:10:41-07:00",
            "Action": "CREATE",
            "State": "FAILED",
            "Error": {
                "ErrorCode": "QueryExecutionException",
                "ErrorMessage": "Error received during view SQL validation using
 a connection: [Connection Name: athena-connection | Query Execution Id: d9bb1e6d-
ce26-4b35-8276-8a199af966aa | Timestamp: Thu Jul 11 18:10:
41 UTC 2024]: Athena validation FAILED: {ErrorCategory: 2,ErrorType: 1301,Retryable:
 false,ErrorMessage: line 1:118: Schema 'gdc--view-playground-db' does not exist}"
            },
            "Details": {
                "RequestedChange": {
                    "Name": "view-athena-redshift-70",
                    "DatabaseName": "async-view-test-db",
                    "Description": "This is an atomic operation",
                    "Retention": 0,
                    "StorageDescriptor": {
                        "Columns": [
                            {
                                "Name": "col1",
                                "Type": "int"
                            },
                            {
                                "Name": "col2",
                                "Type": "string"
                            },
                            {
                                "Name": "col3",
                                "Type": "double"
                            }
                        ],
                        "Compressed": false,
                        "NumberOfBuckets": 0,
                        "SortColumns": [],
                        "StoredAsSubDirectories": false
                    },
                    "TableType": "VIRTUAL_VIEW",
                    "IsRegisteredWithLakeFormation": false,
                    "CatalogId": "123456789012",
                    "IsRowFilteringEnabled": false,
                    "VersionId": "-1",
```

```
                    "DatabaseId": "<databaseID>",
                    "ViewDefinition": {
                        "IsProtected": true,
                        "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
                        "SubObjects": [
                            "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1"
                        ],
                        "Representations": [
                            {
                                "Dialect": "ATHENA",
                                "DialectVersion": "3",
                                "ViewOriginalText": "SELECT * FROM \"gdc--view-
playground-db\".\"table_1\"",
                                "IsStale": false
                            },
                            {
                                "Dialect": "REDSHIFT",
                                "DialectVersion": "1.0",
                                "ViewOriginalText": "SELECT * FROM \"gdc-view-
playground-external-schema\".\"table_1\";",
                                "IsStale": false
                            }
                        ]
                    },
                    "IsMultiDialectView": true
                },
                "ViewValidations": [
                    {
                        "Dialect": "ATHENA",
                        "DialectVersion": "3",
                        "ViewValidationText": "SELECT * FROM \"gdc--view-playground-db
\".\"table_1\"",
                        "UpdateTime": "2024-07-11T11:10:41-07:00",
                        "State": "FAILED",
                        "Error": {
                            "ErrorCode": "QueryExecutionException",
                            "ErrorMessage": "Error received during view SQL validation
 using a connection: [Connection Name: athena-connection | Query Execution Id:
 d9bb1e6d-ce26-4b35-8276-8a199af966aa | Timestamp: Thu J
ul 11 18:10:41 UTC 2024]: Athena validation FAILED: {ErrorCategory: 2,ErrorType:
 1301,Retryable: false,ErrorMessage: line 1:118: Schema 'gdc--view-playground-db' does
 not exist}"
                        }
```

```
                    },
                    {
                        "Dialect": "REDSHIFT",
                        "DialectVersion": "1.0",
                        "ViewValidationText": "SELECT * FROM \"gdc-view-playground-
external-schema\".\"table_1\";",
                        "UpdateTime": "2024-07-11T11:10:41-07:00",
                        "State": "SUCCESS"
                    }
                ]
            }
        }
    }
}
```

Cenário 4: descritores de armazenamento incompatíveis

O SQL fornecido para o dialeto do Athena seleciona `col1` e `col2`, enquanto o SQL para o Redshift seleciona somente `col1`. Isso provoca um erro de incompatibilidade do descritor de armazenamento.

Solicitação `GetTable`:

```
{
    "CatalogId": "123456789012",
    "DatabaseName": "async-view-test-db",
    "TableInput": {
        "Name": "view-athena-redshift-71",
        "Description": "This is an atomic operation",
        "StorageDescriptor": {
            "Columns": [
                { "Name": "col1", "Type": "int" },
                { "Name": "col2", "Type": "string" },
                { "Name": "col3", "Type": "double" }
            ]
        },
        "ViewDefinition": {
            "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
            "SubObjects": [ "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1" ],
            "Representations": [
                {
                    "Dialect": "ATHENA",
```

```
                    "DialectVersion": "3",
                    "ViewOriginalText": "SELECT col1, col2 FROM \"gdc-view-playground-
db\".\"table_1\"",
                    "ValidationConnection": "athena-connection"
                },
                {
                    "Dialect": "REDSHIFT",
                    "DialectVersion": "1.0",
                    "ViewOriginalText": "SELECT col1 FROM \"gdc-view-playground-
external-schema\".\"table_1\";",
                    "ValidationConnection": "redshift-connection"
                }
            ]
        }
    }
}
```

Resposta `GetTable`:

```
IncludeStatusDetails = FALSE

{
    "Table": {
        "Name": "view-athena-redshift-71",
        "DatabaseName": "async-view-test-db",
        "Description": "",
        "CreateTime": "2024-07-11T11:22:02-07:00",
        "UpdateTime": "2024-07-11T11:22:02-07:00",
        "Retention": 0,
        "ViewOriginalText": "",
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "IsMultiDialectView": false,
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:22:02-07:00",
```

```
            "UpdateTime": "2024-07-11T11:23:19-07:00",
            "Action": "CREATE",
            "State": "FAILED"
        }
    }
}

IncludeStatusDetails = TRUE

{
    "Table": {
        "Name": "view-athena-redshift-71",
        "DatabaseName": "async-view-test-db",
        "Description": "",
        "CreateTime": "2024-07-11T11:22:02-07:00",
        "UpdateTime": "2024-07-11T11:22:02-07:00",
        "Retention": 0,
        "ViewOriginalText": "",
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "IsMultiDialectView": false,
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:22:02-07:00",
            "UpdateTime": "2024-07-11T11:23:19-07:00",
            "Action": "CREATE",
            "State": "FAILED",
            "Error": {
                "ErrorCode": "InvalidInputException",
                "ErrorMessage": "Engine and existing storage descriptor mismatch"
            },
            "Details": {
                "RequestedChange": {
                    "Name": "view-athena-redshift-71",
                    "DatabaseName": "async-view-test-db",
                    "Description": "This is an atomic operation",
                    "Retention": 0,
```

```
                    "StorageDescriptor": {
                        "Columns": [
                            {
                                "Name": "col1",
                                "Type": "int"
                            },
                            {
                                "Name": "col2",
                                "Type": "string"
                            },
                            {
                                "Name": "col3",
                                "Type": "double"
                            }
                        ],
                        "Compressed": false,
                        "NumberOfBuckets": 0,
                        "SortColumns": [],
                        "StoredAsSubDirectories": false
                    },
                    "TableType": "VIRTUAL_VIEW",
                    "IsRegisteredWithLakeFormation": false,
                    "CatalogId": "123456789012",
                    "IsRowFilteringEnabled": false,
                    "VersionId": "-1",
                    "DatabaseId": "<databaseID>",
                    "ViewDefinition": {
                        "IsProtected": true,
                        "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
                        "SubObjects": [
                            "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1"
                        ],
                        "Representations": [
                            {
                                "Dialect": "ATHENA",
                                "DialectVersion": "3",
                                "ViewOriginalText": "SELECT col1, col2 FROM \"gdc-view-
playground-db\".\"table_1\"",
                                "IsStale": false
                            },
                            {
                                "Dialect": "REDSHIFT",
                                "DialectVersion": "1.0",
```

```
                                "ViewOriginalText": "SELECT col1 FROM \"gdc-view-
playground-external-schema\".\"table_1\";",
                                "IsStale": false
                            }
                        ]
                    },
                    "IsMultiDialectView": true
                },
                "ViewValidations": [
                    {
                        "Dialect": "ATHENA",
                        "DialectVersion": "3",
                        "ViewValidationText": "SELECT col1, col2 FROM \"gdc-view-
playground-db\".\"table_1\"",
                        "UpdateTime": "2024-07-11T11:23:19-07:00",
                        "State": "FAILED",
                        "Error": {
                            "ErrorCode": "InvalidInputException",
                            "ErrorMessage": "Engine and existing storage descriptor
 mismatch"
                        }
                    },
                    {
                        "Dialect": "REDSHIFT",
                        "DialectVersion": "1.0",
                        "ViewValidationText": "SELECT col1 FROM \"gdc-view-playground-
external-schema\".\"table_1\";",
                        "UpdateTime": "2024-07-11T11:22:49-07:00",
                        "State": "FAILED",
                        "Error": {
                            "ErrorCode": "InvalidInputException",
                            "ErrorMessage": "Engine and existing storage descriptor
 mismatch"
                        }
                    }
                ]
            }
        }
    }
}
```

Conceder permissões nas visualizações do catálogo de dados

Depois de criar visualizações no AWS Glue Data Catalog, você pode conceder permissões de data lake sobre visualizações a diretores de todas Contas da AWS as organizações e unidades organizacionais. Você pode conceder permissões usando LF-Tags ou o método de recurso nomeado. Para obter mais informações sobre os recursos de tag, consulte Controle de acesso baseado em tags do Lake Formation. Para obter mais informações sobre como conceder permissões a visualizações diretamente, consulte Conceder permissões em visualizações usando o método de recurso nomeado.

Visões materializadas

Tópicos

- Diferenciando visualizações materializadas de outros tipos de visualização
- Casos de uso
- Principais conceitos
- Permissões para visões materializadas
- Criação e gerenciamento de visualizações materializadas
- Armazenamento e acesso aos dados
- Integrar ao AWS Lake Formation permissões
- Monitoramento e depuração
- Gerenciando trabalhos de atualização
- Monitoramento e solução de problemas
- Considerações e limitações

No Catálogo de AWS Glue Dados, uma visualização materializada é uma tabela gerenciada que armazena o resultado pré-computado de uma consulta SQL no formato Apache Iceberg. Diferentemente das visualizações padrão do Catálogo de Dados, que executam a consulta sempre que são acessadas, as visualizações materializadas armazenam fisicamente os resultados da consulta e os atualizam à medida que as tabelas de origem subjacentes mudam. Você pode criar visualizações materializadas usando o Apache Spark versão 3.5.6+ no Amazon Athena, Amazon EMR ou. AWS Glue

As visualizações materializadas fazem referência às tabelas do Apache Iceberg registradas no catálogo de AWS Glue dados, com dados pré-computados armazenados como tabelas do Apache

Iceberg em buckets do Amazon S3 Tables ou buckets de uso geral do Amazon S3, tornando-os acessíveis a partir de vários mecanismos de consulta, incluindo Amazon Athena, Amazon Redshift e mecanismos de terceiros. Iceberg-compatible

Diferenciando visualizações materializadas de outros tipos de visualização

As visualizações materializadas diferem das visualizações do catálogo de AWS Glue dados, das visualizações do Apache Spark e das visualizações do Amazon Athena de maneiras fundamentais. Embora as visualizações do Catálogo de Dados sejam tabelas virtuais que executam a definição da consulta SQL sempre que são acessadas, as visualizações materializadas armazenam fisicamente os resultados pré-computados da consulta. Isso elimina a computação redundante e melhora significativamente o desempenho da consulta para transformações complexas acessadas com frequência.

As visualizações materializadas também diferem dos pipelines tradicionais de transformação de dados criados com AWS Glue ETL ou tarefas personalizadas do Spark. Em vez de escrever código personalizado para lidar com a detecção de alterações, atualizações incrementais e orquestração do fluxo de trabalho, você define visualizações materializadas usando a sintaxe SQL padrão. O Catálogo AWS Glue de Dados monitora automaticamente as tabelas de origem, detecta alterações e atualiza as visualizações materializadas usando uma infraestrutura computacional totalmente gerenciada.

Casos de uso

Veja a seguir casos de uso importantes para visualizações materializadas:

- Acelere consultas analíticas complexas — crie visualizações materializadas que pré-computam junções, agregações e funções de janela caras. Os mecanismos do Spark reescrevem automaticamente as consultas subsequentes para usar os resultados pré-computados, reduzindo a latência da consulta e os custos de computação.
- Simplifique os pipelines de transformação de dados — substitua tarefas complexas de ETL que lidam com detecção de alterações, atualizações incrementais e orquestração de fluxo de trabalho por definições simples de visualização materializada. SQL-based O Catálogo AWS Glue de Dados gerencia toda a complexidade operacional automaticamente.
- Habilite análises de autoatendimento com acesso controlado aos dados — Crie visualizações materializadas com curadoria que transformam dados brutos em conjuntos de dados prontos para uso comercial. Conceda aos usuários acesso a visualizações materializadas sem expor as tabelas de origem subjacentes, simplificando o gerenciamento da segurança e fortalecendo a análise de autoatendimento.

- Otimize a engenharia de recursos para aprendizado de máquina — defina visualizações materializadas que implementam transformações de recursos para modelos de ML. O recurso de atualização automática garante que os repositórios de recursos permaneçam atualizados à medida que os dados de origem evoluem, enquanto a atualização incremental minimiza os custos de computação.
- Implemente um compartilhamento eficiente de dados — crie visualizações materializadas que filtram e transformam dados para consumidores específicos. Compartilhe visualizações materializadas entre contas e regiões usando AWS Lake Formation, eliminando a necessidade de duplicação de dados e mantendo a governança centralizada.

Principais conceitos

Atualização automática

A atualização automática é um recurso que monitora continuamente suas tabelas de origem e atualiza as visualizações materializadas de acordo com um cronograma definido por você. Ao criar uma visualização materializada, você pode especificar uma frequência de atualização usando uma programação baseada em tempo com intervalos de até uma hora. O Catálogo de AWS Glue Dados usa a infraestrutura computacional gerenciada do Spark para executar operações de atualização em segundo plano, lidando de forma transparente com todos os aspectos da detecção de alterações e atualizações incrementais.

Quando os dados de origem mudam entre os intervalos de atualização, a visualização materializada fica temporariamente obsoleta. As consultas que acessam diretamente a visualização materializada podem retornar resultados desatualizados até que a próxima atualização programada seja concluída. Para cenários que exigem acesso imediato aos dados mais atuais, você pode executar uma atualização manual usando o comando `REFRESH MATERIALIZED VIEW` SQL.

Atualização incremental

A atualização incremental é uma técnica de otimização que processa somente os dados que foram alterados nas tabelas de origem desde a última atualização, em vez de recomputar toda a visualização materializada. O Catálogo de AWS Glue Dados aproveita a camada de metadados do Apache Iceberg para rastrear com eficiência as alterações nas tabelas de origem e determinar quais partes da visualização materializada precisam de atualizações.

Essa abordagem reduz significativamente os custos de computação e a duração da atualização em comparação com as operações de atualização completa, especialmente para grandes conjuntos

de dados em que apenas uma pequena porcentagem dos dados é alterada entre os ciclos de atualização. O mecanismo de atualização incremental opera automaticamente; você não precisa escrever uma lógica personalizada para detectar ou processar dados alterados.

Reescrita automática de consultas

A reescrita automática de consultas é um recurso de otimização de consultas disponível nos mecanismos do Spark no Amazon Athena, Amazon EMR e. AWS Glue Quando você executa uma consulta em tabelas base, o otimizador do Spark analisa seu plano de consulta e determina automaticamente se as visualizações materializadas disponíveis podem satisfazer a consulta com mais eficiência. Se existir uma visualização materializada adequada, o otimizador reescreve a consulta de forma transparente para usar os resultados pré-computados em vez de processar as tabelas base.

Essa otimização ocorre sem exigir nenhuma alteração no código do aplicativo ou nas instruções de consulta. O otimizador Spark garante que a reescrita automática de consultas só se aplique quando a visualização materializada estiver atualizada e puder produzir resultados precisos. Se uma visualização materializada estiver obsoleta ou não corresponder totalmente aos requisitos de consulta, o otimizador executa o plano de consulta original em relação às tabelas base, priorizando a correção em detrimento do desempenho.

Exibir função do definidor

Uma visualização materializada opera com base nas permissões da função do IAM que a criou, conhecida como função definidora de visualizações. A função definidora deve ter acesso de leitura a todas as tabelas base referenciadas na definição da visualização materializada e criar permissões de tabela no banco de dados de destino. Quando o Catálogo de AWS Glue Dados atualiza uma exibição materializada, ele assume a função de definidor de acessar tabelas de origem e gravar resultados atualizados.

Esse modelo de segurança permite que você conceda aos usuários acesso às visualizações materializadas sem conceder a eles permissões diretas nas tabelas de origem subjacentes. Se a função do definidor de visualizações perder o acesso a qualquer tabela base, as operações de atualização subsequentes falharão até que as permissões sejam restauradas.

Permissões para visões materializadas

Para criar e gerenciar visualizações materializadas, você deve configurar AWS Lake Formation as permissões. O perfil do IAM que cria a visão materializada (o perfil definidor) requer permissões específicas nas tabelas de origem e nos bancos de dados de destino.

Permissões necessárias para o perfil de definidor

O perfil deve ter as seguintes permissões:

- Nas tabelas de origem: permissões SELECT ou ALL sem filtros de linha, coluna ou célula
- No banco de dados de destino: permissão CREATE_TABLE
- No catálogo de AWS Glue dados — GetTable e nas permissões CreateTable da API

Quando você cria uma visão materializada, o ARN do perfil definidor é armazenado na definição da visão. O Catálogo AWS Glue de Dados assume essa função ao executar operações de atualização automática. Se o perfil do definidor perder o acesso às tabelas de origem, as operações de atualização falharão até que as permissões sejam restauradas.

Permissões do IAM para AWS Glue jobs

A AWS Glue função do IAM do seu trabalho exige as seguintes permissões:

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "glue:GetCatalog",
                "glue:GetCatalogs",
                "glue:GetTable",
                "glue:GetTables",
                "glue:CreateTable",
                "glue:UpdateTable",
                "glue:DeleteTable",
                "glue:GetDatabase",
                "glue:GetDatabases",
                "cloudwatch:PutMetricData"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:GetObject",
                "s3:PutObject",
                "s3:DeleteObject"
```

```
            ],
            "Resource": [
                "arn:aws:s3:::amzn-s3-demo-bucket/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::amzn-s3-demo-bucket"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "logs:CreateLogGroup",
                "logs:CreateLogStream",
                "logs:PutLogEvents"
            ],
            "Resource": [
                "arn:aws:logs:*:*:*:/aws-glue/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:GetDataAccess"
            ],
            "Resource": "*"
        }
    ]
}
```

A função que você usa para a atualização automática do Materialized View deve ter a PassRole permissão iam: na função.

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

```
      "Effect": "Allow",
      "Action": [
        "iam:PassRole"
      ],
      "Resource": [
        "arn:aws:iam::111122223333:role/materialized-view-role-name"
      ]
    }
  ]
}
```

Para que o Glue atualize automaticamente a visão materializada para você, o perfil também deve ter a política de confiança a seguir que permite que o serviço assuma o perfil.

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "glue.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

Se a visão materializada estiver armazenada em buckets do Tabelas do S3, também precisará adicionar a seguinte permissão ao perfil.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3tables:PutTableMaintenanceConfiguration"
      ],
      "Resource": "arn:aws:s3tables:*:123456789012:*"
```

```
    }
  ]
}
```

Conceder acesso às visões materializadas

Para conceder a outros usuários acesso para consultar uma visualização materializada, use AWS Lake Formation para conceder a permissão SELECT na tabela de visualização materializada. Os usuários podem consultar a visão materializada sem precisar de acesso direto às tabelas de origem subjacentes.

Para obter informações detalhadas sobre como configurar as permissões do Lake Formation, consulte Conceder e revogar permissões nos recursos do Catálogo de Dados no Guia do AWS Lake Formation Desenvolvedor.

Criação e gerenciamento de visualizações materializadas

Você cria visualizações materializadas usando a `CREATE MATERIALIZED VIEW` instrução SQL nos mecanismos do Spark. A definição da exibição especifica a consulta SQL que define a lógica de transformação, o nome do banco de dados e da tabela de destino e a configuração de atualização opcional. Você pode definir transformações complexas, incluindo agregações, junções em várias tabelas, filtros e funções de janela.

```
CREATE MATERIALIZED VIEW sales_summary
AS
SELECT
    region,
    product_category,
    SUM(sales_amount) as total_sales,
    COUNT(DISTINCT customer_id) as unique_customers
FROM sales_transactions
WHERE transaction_date >= current_date - interval '90' day
GROUP BY region, product_category;
```

Para configurar a atualização automática, inclua a agenda de atualização na sua definição de exibição:

```
CREATE MATERIALIZED VIEW sales_summary
SCHEDULE REFRESH EVERY 1 HOUR
```

```
AS
SELECT region, product_category, SUM(sales_amount) as total_sales
FROM sales_transactions
GROUP BY region, product_category;
```

Você pode atualizar manualmente uma visualização materializada a qualquer momento usando o `REFRESH MATERIALIZED VIEW` comando:

```
REFRESH MATERIALIZED VIEW sales_summary;
```

Para modificar o cronograma de atualização de uma visualização materializada existente, use a `ALTER MATERIALIZED VIEW` instrução:

```
ALTER MATERIALIZED VIEW sales_summary
ADD SCHEDULE REFRESH EVERY 2 HOURS;
```

Visões materializadas aninhadas

Você pode criar visualizações materializadas que referenciem outras visualizações materializadas como tabelas base, permitindo transformações de dados em vários estágios. Quando você cria visualizações materializadas aninhadas, o Catálogo de AWS Glue Dados rastreia dependências e propaga automaticamente as atualizações por meio da hierarquia de visualizações materializadas. Quando uma visualização materializada básica é atualizada, todas as visualizações materializadas posteriores que dependem dela são atualizadas adequadamente.

Esse recurso permite que você decomponha transformações complexas em estágios lógicos, melhorando a capacidade de manutenção e permitindo a atualização seletiva das camadas de transformação com base em seus requisitos de atualização de dados.

Armazenamento e acesso aos dados

As visualizações materializadas armazenam resultados pré-computados como tabelas Apache Iceberg em buckets S3 Tables ou buckets S3 de uso geral em sua conta. AWS O Catálogo AWS Glue de Dados gerencia todos os aspectos da manutenção da tabela Iceberg, incluindo compactação e retenção de instantâneos, por meio dos recursos de otimização automatizada do S3 Tables.

Como as visualizações materializadas são armazenadas como tabelas Iceberg, você pode lê-las diretamente de qualquer Iceberg-compatible mecanismo, incluindo Amazon Athena, Amazon

Redshift e plataformas de análise de terceiros. Essa acessibilidade multimecanismo garante que seus dados pré-computados permaneçam acessíveis em todo o seu ecossistema de análise sem duplicação de dados ou conversão de formato.

Integrar ao AWS Lake Formation permissões

Você pode usar AWS Lake Formation para gerenciar permissões refinadas em visualizações materializadas. O criador da visualização se torna automaticamente o proprietário da visualização materializada e pode conceder permissões a outros usuários ou funções usando o método AWS Lake Formation de recurso nomeado ou LF-Tags.

Quando você concede `SELECT` permissão a um usuário em uma visualização materializada, ele pode consultar os resultados pré-computados sem precisar acessar as tabelas de origem subjacentes. Esse modelo de segurança simplifica o gerenciamento do acesso aos dados e permite que você implemente o princípio do menor privilégio, fornecendo aos usuários acesso somente às transformações de dados específicas de que precisam.

Você pode compartilhar visualizações materializadas entre AWS contas, AWS organizações e unidades organizacionais usando os recursos de compartilhamento entre contas AWS Lake Formation da. Você também pode acessar visualizações materializadas em todas AWS as regiões usando links de recursos, permitindo a governança centralizada de dados com acesso distribuído aos dados.

Monitoramento e depuração

O Catálogo AWS Glue de Dados publica todas as operações de atualização de visualizações materializadas e métricas associadas na Amazon. CloudWatch Você pode monitorar a hora de início, a hora de término, a duração, o volume de dados processados e o status da atualização por meio CloudWatch de métricas. Quando as operações de atualização falham, as mensagens de erro e as informações de diagnóstico são capturadas nos CloudWatch registros.

Você pode configurar CloudWatch alarmes para receber notificações quando os trabalhos de atualização excederem a duração esperada ou falharem repetidamente. O Catálogo de AWS Glue Dados também publica eventos de alteração para execuções de atualização bem-sucedidas e malsucedidas, permitindo que você integre operações de visualização materializada em uma automação mais ampla do fluxo de trabalho.

Para verificar o status atual de uma visualização materializada, use o comando `DESCRIBE MATERIALIZED VIEW` SQL, que retorna metadados, incluindo status de inatividade, data e hora da última atualização e configuração do cronograma de atualização.

Gerenciando trabalhos de atualização

Iniciando uma atualização manual

Acione uma atualização imediata fora do intervalo programado.

Permissão necessária: AWS as credenciais usadas para fazer a chamada à API devem ter `glue:GetTable` permissão para a visualização materializada.

Para o catálogo de tabelas do S3:

```
aws glue start-materialized-view-refresh-task-run \
    --catalog-id <ACCOUNT_ID>:s3tablescatalog/<CATALOG_NAME> \
    --database-name <DATABASE_NAME> \
    --table-name <MV_TABLE_NAME>
```

Para o Root Catalog:

```
aws glue start-materialized-view-refresh-task-run \
    --catalog-id <ACCOUNT_ID> \
    --database-name <DATABASE_NAME> \
    --table-name <MV_TABLE_NAME>
```

Verificando o status da atualização

Obtenha o status de um trabalho de atualização específico:

```
aws glue get-materialized-view-refresh-task-run \
    --catalog-id <CATALOG_ID> \
    --materialized-view-refresh-task-run-id <TASK_RUN_ID>
```

Histórico de atualização da lista

Visualize todas as tarefas de atualização de uma visão materializada:

```
aws glue list-materialized-view-refresh-task-runs \
    --catalog-id <CATALOG_ID> \
    --database-name <DATABASE_NAME> \
    --table-name <MV_TABLE_NAME>
```

> **Note**
>
> Use `<ACCOUNT_ID>:s3tablescatalog/<CATALOG_NAME>` para tabelas S3 ou `<ACCOUNT_ID>` para catálogo raiz.

Interrompendo uma atualização em execução

Cancelar um trabalho de atualização em andamento:

```
aws glue stop-materialized-view-refresh-task-run \
    --catalog-id <CATALOG_ID> \
    --database-name <DATABASE_NAME> \
    --table-name <MV_TABLE_NAME>
```

Monitoramento e solução de problemas

Há três maneiras de monitorar trabalhos de atualização de visualizações materializadas:

CloudWatch Métricas

Visualize métricas agregadas para todas as suas tarefas de atualização de visualização materializada em: CloudWatch

Métricas disponíveis:

- AWS/Glue namespace com dimensões:
  - CatalogId: Seu identificador de catálogo
  - DatabaseName: Banco de dados contendo a visualização materializada
  - TableName: Nome da visualização materializada
  - TaskType: Defina como "MaterializedViewRefresh”

Visualizando no console:

1. Navegue até CloudWatch Console → Métricas
2. Selecione o AWS namespace /Glue
3. Filtrar por dimensões: CatalogId, DatabaseName, TableName, TaskType

4. Veja métricas de sucesso, fracasso e duração do trabalho

Exemplo de consulta de CloudWatch métricas:

```
{AWS/Glue,CatalogId,DatabaseName,TableName,TaskType} MaterializedViewRefresh
```

Usando AWS CLI:

```
aws cloudwatch get-metric-statistics \
    --namespace AWS/Glue \
    --metric-name <MetricName> \
    --dimensions Name=CatalogId,Value=<CATALOG_ID> \
                 Name=DatabaseName,Value=<DATABASE_NAME> \
                 Name=TableName,Value=<TABLE_NAME> \
                 Name=TaskType,Value=MaterializedViewRefresh \
    --start-time <START_TIME> \
    --end-time <END_TIME> \
    --period 3600 \
    --statistics Sum \
    --region <REGION>
```

CloudWatch Registros

Visualize registros de execução detalhados para execuções individuais de tarefas de atualização:

Grupo de registros: `/aws-glue/materialized-views/<task_run_id>`

Onde `<task_run_id>` está um UUID (por exemplo, abc12345-def6-7890-ghij-klmnopqrstuv).

Visualizando registros:

```
# List log streams for a task run
aws logs describe-log-streams \
    --log-group-name /aws-glue/materialized-views/<TASK_RUN_ID> \
    --region <REGION>

# Get log events
aws logs get-log-events \
    --log-group-name /aws-glue/materialized-views/<TASK_RUN_ID> \
    --log-stream-name <LOG_STREAM_NAME> \
    --region <REGION>
```

No CloudWatch console:

1. Navegue até CloudWatch → Grupos de registros
2. Pesquise por /aws- -views/ glue/materialized
3. Selecione o grupo de registros com seu ID de execução da tarefa
4. Visualize registros detalhados de execução, erros e resultados de tarefas do Spark

Notificações

Inscreva-se em eventos para receber notificações em tempo real sobre mudanças no estado da tarefa de atualização:

Tipos de eventos disponíveis:

- Tarefa de atualização do Glue Materialized View iniciada
- A tarefa de atualização do Glue Materialized View foi bem-sucedida
- Falha na tarefa de atualização do Glue Materialized View
- Falha na Auto-Refresh invocação do Glue Materialized View

Criando uma regra:

```
aws events put-rule \
    --name materialized-view-refresh-notifications \
    --event-pattern '{
        "source": ["aws.glue"],
        "detail-type": [
            "Glue Materialized View Refresh Task Started",
            "Glue Materialized View Refresh Task Succeeded",
            "Glue Materialized View Refresh Task Failed",
            "Glue Materialized View Auto-Refresh Invocation Failure"
        ]
    }' \
    --region <REGION>
```

Adicionando um alvo (por exemplo, tópico do SNS):

```
aws events put-targets \
    --rule materialized-view-refresh-notifications \
```

```
    --targets "Id"="1","Arn"="arn:aws:sns:<REGION>:<ACCOUNT_ID>:<TOPIC_NAME>" \
    --region <REGION>
```

Visualizando o status de atualização

Verifique o status de seus trabalhos de atualização de visualização materializada usando a AWS Glue API:

```
aws glue get-materialized-view-refresh-task-run \
    --catalog-id <CATALOG_ID> \
    --materialized-view-refresh-task-run-id <TASK_RUN_ID> \
    --region <REGION>
```

Ou liste todas as execuções de atualização recentes:

```
aws glue list-materialized-view-refresh-task-runs \
    --catalog-id <CATALOG_ID> \
    --database-name <DATABASE_NAME> \
    --table-name <MV_TABLE_NAME> \
    --region <REGION>
```

Isso mostra:

- Hora da última atualização
- Status de atualização (BEM-SUCEDIDA, FALHA, EM EXECUÇÃO, PARADA)
- ID de execução da tarefa
- Mensagens de erro (se falharem)

Estados comuns de atualização:

- EM EXECUÇÃO: O trabalho de atualização está sendo executado no momento
- BEM-SUCEDIDO: Atualização concluída com sucesso
- FALHA: A atualização encontrou um erro
- PARADO: a atualização foi cancelada manualmente

Solução de problemas de atualizações com falha:

Se uma atualização falhar, verifique:

1. Permissões do IAM: garanta que a função definidora tenha acesso a todas as tabelas base e ao local da visualização materializada
2. Disponibilidade da tabela base: verifique se todas as tabelas referenciadas existem e estão acessíveis
3. Validade da consulta: confirme se a consulta SQL é válida para o dialeto SQL do Spark
4. Limites de recursos: verifique se você atingiu os limites de atualização simultânea da sua conta

Use a GetMaterializedViewRefreshTaskRun API para recuperar mensagens de erro detalhadas.

Considerações e limitações

- As visualizações materializadas só podem referenciar tabelas do Apache Iceberg registradas no Catálogo de AWS Glue Dados como tabelas base.
- A criação de visualizações e a reescrita automática de consultas estão disponíveis somente nos mecanismos Spark no Apache Spark versão 3.5.6 e superior no Amazon Athena, Amazon EMR e (versão 5.1). AWS Glue
- Eventualmente, as visualizações materializadas são consistentes com as tabelas base. Durante a janela de atualização, as consultas que acessam diretamente a visualização materializada podem retornar dados desatualizados. Para acesso imediato aos dados atuais, execute uma atualização manual.
- O intervalo mínimo de atualização automática é uma hora. Para casos de uso que exigem atualizações mais frequentes, execute atualizações manuais programaticamente usando o comando. `REFRESH MATERIALIZED VIEW`
- A reescrita da consulta prioriza a correção em detrimento do desempenho. Se uma visualização materializada estiver obsoleta ou não puder atender aos requisitos de consulta com precisão, os mecanismos do Spark executam a consulta original nas tabelas base.

Importação de dados usando fluxos de trabalho no Lake Formation

Com AWS Lake Formation, você pode importar seus dados usando fluxos de trabalho. Um fluxo de trabalho define a fonte de dados e o cronograma para importar dados para o seu data lake. É um contêiner para crawlers do AWS Glue, trabalhos e acionadores usados para orquestrar os processos de carregamento e atualização do data lake.

Tópicos

- Esquemas e fluxos de trabalho no Lake Formation
- Criação de um fluxo de trabalho
- Executar um fluxo de trabalho

Esquemas e fluxos de trabalho no Lake Formation

Um fluxo de trabalho encapsula uma atividade complexa de vários trabalhos de extração, transformação e carregamento (ETL). Os fluxos de trabalho geram AWS Glue rastreadores, trabalhos e gatilhos para orquestrar o carregamento e a atualização dos dados. O Lake Formation executa e rastreia um fluxo de trabalho como uma entidade única. Você pode configurar um fluxo de trabalho para ser executado sob demanda ou de acordo com um cronograma.

> **Note**
>
> O gravador Parquet do Spark não comporta caracteres especiais nos nomes das colunas. Essa é uma limitação técnica do próprio gravador, não um problema de configuração.

Os fluxos de trabalho que você cria no Lake Formation são visíveis no console do AWS Glue como um gráfico acíclico direcionado (DAG). Cada nó do DAG é uma tarefa, um crawler ou um gatilho. Para monitorar o progresso e solucionar problemas, você pode acompanhar o status de cada nó no fluxo de trabalho.

Quando um fluxo de trabalho do Lake Formation é concluído, o usuário que executou o fluxo de trabalho recebe a permissão `SELECT` do Lake Formation nas tabelas do catálogo de dados que o fluxo de trabalho cria.

Você também pode criar fluxos de trabalho no AWS Glue. No entanto, como o Lake Formation permite que você crie um fluxo de trabalho a partir de um esquema, criar fluxos de trabalho é muito mais simples e automatizado no Lake Formation. Lake Formation fornece os seguintes tipos de esquemas:

- Instantâneo do banco de dados: carrega ou recarrega dados de todas as tabelas no data lake a partir de uma fonte JDBC. Você pode excluir alguns dados da fonte com base em um padrão de exclusão.
- Banco de dados incremental: carrega somente novos dados no data lake a partir de uma fonte JDBC, com base em marcadores previamente definidos. Você especifica as tabelas individuais no banco de dados de origem do JDBC a serem incluídas. Para cada tabela, você escolhe as colunas

dos favoritos e a ordem de classificação dos favoritos para acompanhar os dados que foram carregados anteriormente. Na primeira vez que você executa um esquema de banco de dados incremental em um conjunto de tabelas, o fluxo de trabalho carrega todos os dados das tabelas e define marcadores para a próxima execução do esquema de banco de dados incremental. Portanto, você pode usar um esquema de banco de dados incremental em vez do esquema de instantâneo do banco de dados para carregar todos os dados, desde que você especifique cada tabela na fonte de dados como um parâmetro.

- Arquivo de log — carrega dados em massa de fontes de arquivos de log AWS CloudTrail, incluindo registros do Elastic Load Balancing e registros do Application Load Balancer.

Use a tabela a seguir para ajudar a decidir se deve usar um snapshot de banco de dados ou um esquema de banco de dados incremental.

Use o instantâneo do banco de dados quando...	Use o banco de dados incremental quando...
• A evolução do esquema é flexível. (As colunas são renomeadas, as colunas anteriores são excluídas e novas colunas são adicionadas em seu lugar.) • É necessária uma consistência completa entre a origem e destino.	• A evolução do esquema é incremental. (Há somente adição sucessiva de colunas.) • Somente novas linhas são adicionadas; as linhas anteriores não são atualizadas.

> **Note**
>
> Os usuários não podem editar plantas e fluxos de trabalho criados pelo Lake Formation.

Criação de um fluxo de trabalho

Antes de começar, verifique se você concedeu as permissões de dados e as permissões de localização de dados necessárias para a função `LakeFormationWorkflowRole`. Isso é para que o fluxo de trabalho possa criar tabelas de metadados no catálogo de dados e gravar dados em locais de destino no Amazon S3. Para obter mais informações, consulte (Opcional) Criar um perfil do IAM para fluxos de trabalho e Visão geral das permissões do Lake Formation.

> **Note**
>
> O Lake Formation usa as operações `GetTemplateInstance`, `GetTemplateInstances` e `InstantiateTemplate` para criar fluxos de trabalho com base em esquemas. Essas operações não estão disponíveis ao público e são usadas apenas internamente para criar recursos em seu nome. Você recebe CloudTrail eventos para criar fluxos de trabalho.

Para criar um fluxo de trabalho a partir de um esquema

1. Abra o AWS Lake Formation console em https://console.aws.amazon.com/lakeformation/. Faça login como administrador do data lake ou como usuário com permissões de engenheiro de dados. Para obter mais informações, consulte Referência de personas e permissões do IAM do Lake Formation.
2. No painel de navegação, selecione esquemas e escolha Usar esquema.
3. Na página Usar um esquema, escolha um quadro para selecionar o tipo de esquema.
4. Em Fonte de importação, especifique a fonte de dados.

   Se você estiver importando de uma fonte JDBC, especifique o seguinte:

   - Conexão de banco de dados: escolha uma conexão na lista. Crie conexões adicionais usando o console do AWS Glue. O nome de usuário JDBC e a senha na conexão determinam os objetos do banco de dados aos quais o fluxo de trabalho tem acesso.
   - Caminho dos dados de origem — insira *<database><schema>*//*<table>*ou*<database>*/*<table>*, dependendo do produto do banco de dados. Oracle Database e MySQL não oferecem suporte a esquema no caminho. Você pode substituir o caractere de porcentagem (%) por *<schema>* ou*<table>*. Por exemplo, para um banco de dados Oracle com um identificador de sistema (SID) `orcl`, informe `orcl/%` para importar todas as tabelas às quais o usuário nomeado na conexão tem acesso.

     > **Important**
     >
     > Este campo diferencia letras maiúsculas de minúsculas. O fluxo de trabalho falhará se houver uma incompatibilidade de maiúsculas e minúsculas em qualquer um dos componentes.

Se você especificar um banco de dados MySQL, o AWS Glue ETL usa o driver JDBC Mysql5 por padrão, portanto, o MySQL8 não é suportado nativamente. Você pode editar o script de trabalho ETL para usar um parâmetro `customJdbcDriverS3Path` conforme descrito em Valores connectionType do JDBC no Guia do desenvolvedor do AWS Glue para usar um driver JDBC diferente que suporte MySQL8.

Se você estiver importando de um arquivo de log, certifique-se de que a função especificada para o fluxo de trabalho (a "função do fluxo de trabalho") tenha as permissões necessárias do IAM para acessar a fonte de dados. Por exemplo, para importar AWS CloudTrail registros, o usuário deve ter as `cloudtrail:LookupEvents` permissões `cloudtrail:DescribeTrails` e para ver a lista de CloudTrail registros ao criar o fluxo de trabalho, e a função do fluxo de trabalho deve ter permissões no CloudTrail local no Amazon S3.

5. Execute um destes procedimentos:

   - Para o tipo de esquema de instantâneo do banco de dados, identifique opcionalmente um subconjunto de dados a serem importados especificando um ou mais padrões de exclusão. Esses padrões de exclusão são Unix-style `glob` padrões. Elas são armazenadas como uma propriedade das tabelas criadas pelo fluxo de trabalho.

     Para obter detalhes sobre os padrões de exclusão disponíveis, consulte Incluir e excluir padrões no Guia do desenvolvedor do AWS Glue .

   - Para o tipo de esquema de banco de dados incremental, especifique os seguintes campos. Adicione uma linha para cada tabela a ser importada.

     Nome da tabela

     Tabela a ser importada. Deve estar em letras minúsculas.

     Teclas de marcadores

     Comma-delimited lista de nomes de colunas que definem as teclas de favoritos. Se estiver em branco, a chave primária será usada para determinar novos dados. As maiúsculas e minúsculas de cada coluna devem corresponder às maiúsculas e minúsculas, conforme definido na fonte de dados.

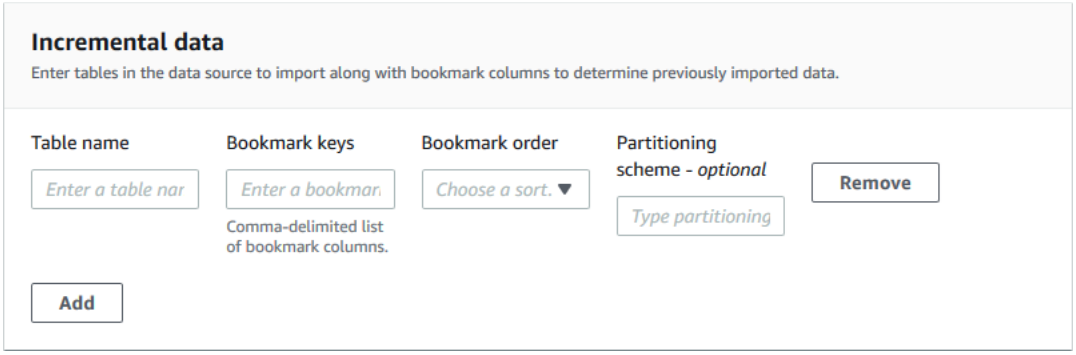
> **Note**
>
> A chave primária se qualifica como a chave de bookmark padrão apenas se estiver aumentando ou diminuindo sequencialmente (sem lacunas). Se você quiser usar a chave primária como chave de marcador e ela tiver lacunas, você deverá nomear a coluna da chave primária como chave de marcador.

Pedido de favoritos

Quando você escolhe Ascendente, as linhas com valores maiores que os marcados são identificadas como novas. Quando você escolhe Descendente, as linhas com valores menores que os valores marcados são identificadas como novas.

Esquema de particionamento

(Opcional) Lista de colunas-chave de particionamento, delimitada por barras (/). Exemplo: `year/month/day`.



Para obter mais informações, consulte Rastreamento de dados processados usando marcadores de trabalho no Guia do desenvolvedor do AWS Glue .

6. Em Importar destino, especifique o banco de dados de destino, a localização de destino do Amazon S3 e o formato dos dados.

   Certifique-se de que a função do fluxo de trabalho tenha as permissões necessárias do Lake Formation no banco de dados e no local de destino do Amazon S3.

   > **Note**
   >
   > Atualmente, os esquemas não oferecem suporte à criptografia de dados no destino.

7. Escolha uma frequência de importação.

   Você pode especificar uma expressão `cron` com a opção Personalizada.

8. Em Opções de importação:

   a. Insira um nome de fluxo de trabalho.

   b. Para a função, escolha a função `LakeFormationWorkflowRole`, que você criou em (Opcional) Criar um perfil do IAM para fluxos de trabalho.

   c. Opcionalmente, especifique um prefixo de tabela. O prefixo é anexado aos nomes das tabelas do catálogo de dados que o fluxo de trabalho cria.

9. Selecione Criar e aguarde até que o console informe que o fluxo de trabalho foi criado com sucesso.

> **Tip**
>
> Você recebeu a seguinte mensagem de erro?
> `User: arn:aws:iam::<account-id>:user/<username> is not authorized to perform: iam:PassRole on resource:arn:aws:iam::<account-id>:role/<rolename>...`
> Nesse caso, verifique se você *<account-id>* substituiu por um número de AWS conta válido em todas as políticas.

> **Consulte também:**
>
> - Esquemas e fluxos de trabalho no Lake Formation

Executar um fluxo de trabalho

Você pode executar um fluxo de trabalho usando o console do Lake Formation, o console do AWS Glue, a interface de linha de comando (AWS CLI) do AWS Glue ou a API.

Para executar um fluxo de trabalho (console do Lake Formation)

1. Abra o AWS Lake Formation console em https://console.aws.amazon.com/lakeformation/. Faça login como administrador do data lake ou como usuário com permissões de engenheiro de

dados. Para obter mais informações, consulte Referência de personas e permissões do IAM do Lake Formation.

2. No painel de navegação, escolha Esquemas.
3. Na página Esquemas, selecione o fluxo de trabalho. Em seguida, no menu Ações, escolha Iniciar.
4. À medida que o fluxo de trabalho é executado, você pode ver o progresso na coluna de Status da última execução. Escolha o botão de atualização ocasionalmente.

   O status vai de EM EXECUÇÃO para Descoberta, para Importação e para CONCLUÍDO.

   Quando o fluxo de trabalho for concluído:

   - O catálogo de dados tem novas tabelas de metadados.
   - Seus dados são ingeridos no data lake.

   Se o fluxo de trabalho falhar, faça o seguinte:

   a. Selecione o fluxo de trabalho. Escolha Ações e Exibir gráfico.

      O fluxo de trabalho é aberto no console do AWS Glue.

   b. Certifique-se de que o fluxo de trabalho esteja selecionado e acesse a guia Histórico.
   c. Em Histórico, selecione a execução mais recente e selecione Exibir informações da execução.
   d. Selecione um trabalho ou crawler com falha no gráfico dinâmico (runtime) e revise a mensagem de erro. Os nós com falha são vermelhos ou amarelos.

Consulte também:

- Esquemas e fluxos de trabalho no Lake Formation

Trazendo seus dados para o AWS Glue Data Catalog

Você pode criar catálogos federados no AWS Glue Data Catalog (Catálogo de dados) e unificar dados entre os data lakes do Amazon S3 e os armazéns de dados do Amazon Redshift. Você também pode integrar dados de seus bancos de dados operacionais Amazon DynamoDB, como fontes de dados de terceiros, como PostgreSQL BigQuery, Google, MySQL, entre outros. O Catálogo de Dados oferece um repositório centralizado de metadados que facilita o gerenciamento e a descoberta de dados em sistemas diferentes.

O Data Catalog se integra a mais de trinta fontes de dados externas por meio de conectores federados. Com essa integração, você pode consultar dados dessas fontes externas sem precisar criar pipelines de dados para ingerir os dados primeiro. AWS

Depois de catalogar os dados externos, você pode usar AWS Lake Formation para gerenciar centralmente as permissões de acesso aos dados no Catálogo de Dados. Os administradores do data lake podem conceder permissões de acesso refinadas a outras entidades principais do IAM (usuários ou perfis) na mesma conta ou em várias contas usando controle de acesso baseado em tag (tags LF) e métodos de recursos nomeados.

Ao usar tags LF, os administradores de dados podem organizar logicamente os recursos com base em atributos como domínio e nível de sensibilidade, simplificando o gerenciamento de permissões e garantindo controles de acesso consistentes em serviços de análise e aprendizado de máquina, incluindo Athena, Amazon EMR ou Redshift Spectrum. AWS Glue

O Data Catalog fornece os seguintes métodos para gerenciar dados e permissões em conjuntos de dados e metastores externos:

- Integre buckets de tabela do Amazon S3 com o catálogo de dados — Você pode publicar e catalogar tabelas do Amazon S3 como objetos do catálogo de dados e registrar o catálogo como um local de dados do Lake Formation a partir do console do Lake Formation ou usando operações de API. AWS Glue
- Traga dados dos armazéns de dados do Amazon Redshift para o AWS Glue Data Catalog — Registre um namespace existente do Amazon Redshift ou um cluster com o Catálogo de Dados e crie um catálogo federado de vários níveis no Catálogo de Dados.

  Você pode acessar seus dados usando qualquer mecanismo de consulta compatível com a especificação OpenAPI do catálogo REST do Apache Iceberg, como Amazon EMR Sem Servidor e Amazon Athena.

- Federe Iceberg REST catálogos remotos ao catálogo de dados — Federe Iceberg REST catálogos remotos ao catálogo de dados e acesse com segurança tabelas remotas armazenadas Iceberg no Amazon S3 usando mecanismos de análise. AWS
- Integre-se ao Catálogo de Dados a partir de fontes de dados externas — Conecte o Catálogo de Dados a fontes de dados externas usando AWS Glue conexões e crie catálogos federados para gerenciar centralmente as permissões de acesso em conjuntos de dados usando o Lake Formation. Não é necessária nenhuma migração de metadados para o Data Catalog.
- Criar catálogos para gerenciar Tabelas do Amazon Redshift no Data Catalog: você pode não ter um cluster produtor do Amazon Redshift ou uma unidade de compartilhamento de dados do Amazon Redshift disponível atualmente, mas deseja criar e gerenciar Tabelas do Amazon Redshift usando o Data Catalog. É possível começar criando um catálogo gerenciado do AWS Glue usando a operação de API `glue:CreateCatalog` ou o console do AWS Lake Formation definindo o tipo de catálogo como `Managed` e `Catalog source` como Redshift.
- Publicar unidades de compartilhamento de dados do Amazon Redshift com o Data Catalog: publique unidades de compartilhamento de dados do Amazon Redshift no Data Catalog e use o Lake Formation para gerenciar centralmente o acesso aos dados das unidades de compartilhamento de dados e restringir o acesso do usuário.

  É possível consultar dados usando o Amazon Redshift Spectrum.
- Conectar o Data Catalog a metastores Hive externos: conecte o Data Catalog a metastores externos para gerenciar permissões de acesso em conjuntos de dados no Amazon S3 usando o Lake Formation. Não é necessária nenhuma migração de metadados para o Data Catalog.
- Integre o Lake Formation com o AWS Data Exchange — O Lake Formation oferece suporte ao licenciamento de acesso aos seus dados por meio AWS Data Exchange de. Se você quiser licenciar seus dados do Lake Formation, consulte O que é o AWS Data Exchange no Guia do usuário do AWS Data Exchange .

Tópicos

- Federação de catálogos para catálogos remotos do Iceberg
- Trazendo dados do Amazon Redshift para o AWS Glue Data Catalog
- Federando em fontes de dados externas no AWS Glue Data Catalog
- Integração de tabelas do Amazon S3 com AWS Glue Data Catalog and AWS Lake Formation
- Alterando os controles de acesso para integração com tabelas do S3
- Criação de um catálogo gerenciado do Amazon Redshift no AWS Glue Data Catalog

- Gerenciamento de permissões para dados em uma unidade de compartilhamento de dados do Amazon Redshift.
- Gerenciamento de permissões em conjuntos de dados que usam repositórios de dados externos

Federação de catálogos para catálogos remotos do Iceberg

A federação de catálogos AWS Glue fornece acesso direto e seguro às tabelas Iceberg, armazenadas no Amazon S3 e catalogadas em catálogos remotos, usando mecanismos de análise. AWS A federação de catálogos sincroniza metadados entre catálogos de dados e catálogos remotos quando você acessa tabelas remotas. Ele é suportado por uma grande variedade de mecanismos de análise, incluindo Amazon Redshift, Amazon EMR, Amazon Athena, mecanismos de terceiros AWS Glue, como o Apache Spark e muito mais.

A federação de catálogos usa AWS Glue Data Catalog para se comunicar com sistemas de catálogos remotos para descobrir tabelas e com o Lake Formation para autorizar o acesso aos dados da tabela no Amazon S3. Quando você consulta uma tabela federada, o Data Catalog descobre as informações mais recentes da tabela no catálogo remoto no momento da consulta, obtendo a localização da tabela no Amazon S3, o esquema atual e as informações de partição. Seu mecanismo de análise (Amazon Athena, Amazon Redshift, Amazon EMR) então usa essas informações para acessar os arquivos de dados do Iceberg diretamente do Amazon S3. O Lake Formation gerencia o acesso às tabelas vendendo credenciais com escopo definido para os dados da tabela armazenados no Amazon S3, permitindo que os mecanismos apliquem permissões refinadas às tabelas federadas.

Características da Federação de Catálogos

Governado usando Lake Formation

Os catálogos federados do Iceberg no Catálogo de Dados são recursos registrados do Lake Formation, permitindo que você conceda permissões refinadas em nível de linha, coluna e célula às tabelas do Iceberg em catálogos federados do Iceberg usando subsídios do Lake Formation. Os catálogos federados do Iceberg e os objetos associados podem ser compartilhados com segurança entre contas. AWS Os catálogos Federated Iceberg também funcionam com o controle de Tag-based acesso do Lake Formation, permitindo que você escale a governança usando tags.

Configurações de rede

A federação de catálogos oferece suporte a conexões diretas com fontes de catálogos remotas usando conectividade HTTPS padrão. Ele também oferece suporte à conectividade por meio da

Amazon VPC quando você deseja manter o isolamento e a conectividade da rede usando o suporte de proxy quando quiser comunicação segura por meio dos firewalls da organização.

Limitações

- Limite de tamanho de metadados da tabela Iceberg — Para catálogos federados do Iceberg, AWS Glue Data Catalog suporta um tamanho máximo de metadados de 5 MB por chamada da API REST. AWS Glue Data Catalog rejeita solicitações para tabelas com metadados que excedem esse limite.

Tópicos

- Federar para Snowflake Catálogo Iceberg
- Federar para Catálogo do Unity do Databricks

Federar para Snowflake Catálogo Iceberg

AWS Glue Data Catalog se federa para Snowflake usar as credenciais OAuth2 de um diretor de serviço. Snowflake Você pode usar essa federação para se conectar Snowflake Horizon e Snowflake Polaris catálogos. Esse mecanismo de autenticação permite que o Catálogo de Dados acesse os metadados de vários objetos (como catálogos, bancos de dados e tabelas) em seu catálogo, com base nos privilégios associados ao principal do serviço. Para garantir o acesso aos objetos certos, é essencial conceder ao responsável pelo serviço as permissões necessárias para ler os metadados desses objetos. Snowflake

Pré-requisitos

Antes de criar um catálogo federado no Catálogo de Dados que seja governado pelo Lake Formation, certifique-se de ter as seguintes permissões:

Seu principal do IAM (usuário ou função) deve ter as seguintes permissões:

- Permissões de formação de lagos —`lakeformation:RegisterResource`, `lakeformation:DescribeResource`
- AWS Glue permissões —`glue:CreateConnection`,`glue:CreateCatalog`,`glue:GetConnection`, `glue:PassConnection`

- Permissões do Secrets Manager —`secretsmanager:CreateSecret`, `secretsmanager:GetSecretValue`
- Permissões do IAM —`iam:CreateRole`,`iam:AttachRolePolicy`, `iam:PassRole`

Você deve ser administrador de data lake da Lake Formation ou ter `CREATE_CATALOG` permissão no Catálogo de Dados

Criar catálogo federado

Usando o console

1. Faça login no console e abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.
2. Escolha a AWS região preferida no seletor de regiões.
3. No painel de navegação à esquerda, selecione Catálogos.
4. Escolha Criar catálogo para abrir o fluxo de trabalho Criar catálogo.
5. Na etapa Escolher fonte de dados, Snowflake selecione uma das opções disponíveis.
6. Na etapa Definir detalhes do catálogo, você fornece três informações: detalhes do catálogo, detalhes da conexão e detalhes do registro.
7. No contêiner de detalhes do catálogo, forneça um nome exclusivo para seu catálogo AWS Glue federado e insira o nome do Snowflake catálogo existente.
8. No contêiner de detalhes das conexões, você pode escolher entre uma conexão existente à qual tenha acesso ou fornecer a configuração para criar um novo conector.
9. As novas configurações de conexão incluem:
   - Nome da conexão — Um nome exclusivo do objeto de AWS Glue conexão.
   - URL da instância — A URL do endpoint da sua Snowflake conta existente.
   - Autenticação — Especifique a configuração de autenticação AWS Glue usada para se conectar ao servidor de catálogo remoto. AWS Glue suporta tanto o OAuth2 quanto a autenticação personalizada.
   - URL do token: especifique o URL do provedor de identidades do catálogo remoto.
   - ID do cliente OAuth2 — (opcional) Especifique a ID do cliente da credencial OAuth2 associada ao seu catálogo remoto. Se você não fornecer um ID de cliente, a solicitação de token será autenticada usando somente o segredo do cliente armazenado nele AWS Secrets Manager.

- Segredo — Armazene e use o segredo do cliente OAuth2 usando AWS Secrets Manager ou insira o valor secreto na caixa de texto. Quando você insere o segredo manualmente no console, AWS Glue cria o segredo em seu nome.
- Escopo do URL do token: especifique o escopo do OAuth para autenticação.
- Filtro de caixa do catálogo — Escolha se deseja trazer objetos em letras minúsculas ou maiúsculas do seu catálogo remoto para o Catálogo de Dados.

10. Crie uma função do IAM que AWS Glue os diretores de serviço do Lake Formation possam usar para acessar locais secretos nas tabelas remotas do Iceberg AWS Secrets Manager e do Amazon S3, respectivamente. Selecione a função do IAM no menu suspenso de registro. Consulte as etapas 2 e 3 na seção CLI a seguir para obter detalhes da política do IAM.
11. Selecione Testar conexão para testar se suas propriedades de conexão e o acesso à função do IAM estão configurados corretamente.
12. Selecione Avançar para revisar suas configurações.
13. Selecione Criar catálogo na página de revisão.

Usar a CLI

1. Crie um AWS Secrets Manager segredo

   O AWS Glue conector oferece suporte a dois tipos de autenticação: OAuth2 e Custom. Ao usar a opção OAuth2, use AWS Secrets Manager para armazenar o segredo do cliente do principal do Snowflake serviço. Posteriormente, você usará esse segredo ao criar a AWS Glue conexão. Para autenticação personalizada, use AWS Secrets Manager para armazenar e recuperar o token de acesso.

   No exemplo a seguir, substitua*your-snowflake-secret*,*client_secret*, *region* por suas próprias informações.

```
aws secretsmanager create-secret \
--name your-snowflake-secret \
--description "Snowflake secret" \
--secret-string '{
"USER_MANAGED_CLIENT_APPLICATION_CLIENT_SECRET": "client_secret"
}' \
--region region
```

> **Note**
>
> `USER_MANAGED_CLIENT_APPLICATION_CLIENT_SECRET`é uma palavra-chave reservada AWS Glue usada para se referir a um valor secreto do cliente no segredo. Use a mesma palavra-chave ao criar o segredo também no console do Lake Formation.

2. Crie uma função do IAM que dê acesso ao objeto de AWS Glue conexão ao segredo criado na etapa anterior

   O objeto de AWS Glue conexão requer acesso ao AWS Secrets Manager segredo quando você o usa AWS Secrets Manager para armazenar, recuperar e atualizar seu token secreto do OAuth. O objeto de AWS Glue conexão também requer acesso para criar, descrever e usar as interfaces de rede Amazon VPC quando você usa um endpoint Amazon VPC para restringir a conectividade à sua conta. Snowflake

   Crie uma política do IAM e anexe-a a uma função do IAM. Adicione o principal do AWS Glue serviço à política de confiança.

   No exemplo a seguir, substitua *`your-secrets-manager-ARNyour-vpc-id`*, e *`your-subnet-id1`* por suas próprias informações.

   Example Política do IAM

```
{
    "Version": "2012-10-17",
    "Statement": [{
            "Effect": "Allow",
            "Action": [
                "secretsmanager:GetSecretValue",
                "secretsmanager:DescribeSecret",
                "secretsmanager:PutSecretValue"
            ],
            "Resource": [
                "your-secrets-manager-ARN"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "ec2:CreateNetworkInterface",
```

```
                "ec2:DeleteNetworkInterface",
                "ec2:DescribeNetworkInterfaces"
            ],
            "Resource": "*",
            "Condition": {
                "ArnEquals": {
                    "ec2:Vpc": "arn:aws:ec2:region:account-id:vpc/your-vpc-id",
                    "ec2:Subnet": ["arn:aws:ec2:region:account-id:subnet/your-
subnet-id1"]
                }
            }
        }
    ]
}
```

Example Política de confiança

```
{
    "Version": "2012-10-17",
    "Statement": [{
        "Effect": "Allow",
        "Principal": {
            "Service": "glue.amazonaws.com"
        },
        "Action": "sts:AssumeRole"
    }]
}
```

3. Crie uma política de IAM que dê ao Lake Formation acesso de leitura à localização do catálogo no Amazon S3

   Como proprietário do catálogo de um catálogo federado no Data Catalog, você usa o Lake Formation para conceder acesso detalhado a tabelas, acesso refinado (em nível de coluna, linha e célula) e acesso baseado em tags às suas equipes de dados. O Lake Formation usa uma função do IAM que lhe dá acesso às localizações subjacentes do Amazon S3 de suas tabelas remotas do Iceberg. Esse acesso permite que a Lake Formation venda credenciais de acesso com escopo definido para mecanismos de análise que consultam tabelas remotas.

   Crie uma política do IAM e anexe-a a uma função do IAM. Adicione o diretor de serviço da Lake Formation à política de confiança de funções.

No exemplo a seguir, substitua *amzn-s3-demo-bucketN* e *your-kms-key* por suas próprias informações.

Example Política do IAM

```
{
    "Version": "2012-10-17",
    "Statement": [{
            "Effect": "Allow",
            "Action": [
                "s3:GetObject"
            ],
            "Resource": [
                "arn:aws:s3:::amzn-s3-demo-bucket1/*",
                "arn:aws:s3:::amzn-s3-demo-bucket2/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::amzn-s3-demo-bucket1",
                "arn:aws:s3:::amzn-s3-demo-bucket2"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "kms:Decrypt",
                "kms:Encrypt"
            ],
            "Resource": [
                "your-kms-key"
            ]
        }
    ]
}
```

Example Política de confiança

```
{
    "Version": "2012-10-17",
    "Statement": [{
        "Sid": "",
        "Effect": "Allow",
        "Principal": {
            "Service": "lakeformation.amazonaws.com"
        },
        "Action": "sts:AssumeRole"
    }]
}
```

> **Note**
>
> Quando você usa o console do Lake Formation para criar um catálogo federado, o console usa uma única função do IAM com as duas políticas anexadas para concluir a configuração.

4. Criar um objeto de AWS Glue conexão

AWS Glue O conector suporta OAuth2 e métodos de autenticação personalizados. O objeto de conexão do Catálogo de Dados oferece suporte à `CATALOG_CASING_FILTER` configuração que permite que você traga objetos em letras minúsculas ou maiúsculas em seu catálogo remoto para o Catálogo de Dados.

O exemplo a seguir usa a configuração de autenticação do OAuth2 para criar uma AWS Glue conexão. Substitua *`highlighted sections`* por suas informações.

```
aws glue create-connection \
    --connection-input '{
"Name": "your-glue-connection-to-snowflake-account",
"ConnectionType": "SNOWFLAKEICEBERGRESTCATALOG",
"ConnectionProperties": {
    "INSTANCE_URL": "your-snowflake-account-URL",
    "ROLE_ARN": "your-IAM-role-for-secrets-and-VPC-access",
    "CATALOG_CASING_FILTER": "LOWERCASE_ONLY"
},
"AuthenticationConfiguration": {
```

```
    "AuthenticationType": "OAUTH2",
    "OAuth2Properties": {
        "OAuth2GrantType": "CLIENT_CREDENTIALS",
        "TokenUrl": "your-internal-or-external-token-server-url",
        "OAuth2ClientApplication": {
            "UserManagedClientApplicationClientId": "our-client-id"
        },
        "TokenUrlParametersMap": {
            "scope": "all-apis"
        }
    },
    "SecretArn": "arn:aws:secretsmanager:your-aws-region:your-aws-account-
id:secret:snowflake-secret"
}
}'
```

> **Note**
>
> O campo `OAuth2ClientApplication` é opcional. Se você omiti-la, a solicitação de token será autenticada usando somente o segredo do cliente armazenado em AWS Secrets Manager, sem enviar uma ID de cliente.

5. Registre AWS Glue a conexão como um recurso de formação de lago

   Usando o objeto de AWS Glue conexão (criado na Etapa 4) e a função IAM (criada na Etapa 3), agora você pode registrar o objeto de AWS Glue conexão como um recurso gerenciado do Lake Formation.

   Substitua *your-glue-connector-arn* e *your-IAM-role-ARN-having-LF-access* por suas informações.

   ```
   aws lakeformation register-resource \
       --resource-arn your-glue-connector-arn \
       --role-arn your-IAM-role-ARN-having-LF-access \
       --with-federation \
       --with-privileged-access
   ```

6. Crie um catálogo federado no Catálogo de Dados

   Depois de criar um objeto de AWS Glue conexão e registrá-lo no Lake Formation, você pode criar um catálogo federado no Catálogo de Dados:

Forneça ao catálogo federado um nome exclusivo em*your-federated-catalog-name*, faça referência ao catálogo em Snowflake at *catalog-name-in-Snowflake* e insira o nome da conexão criada anteriormente em*your-glue-connection-name*.

```
aws glue create-catalog \
    --name your-federated-catalog-name \
    --catalog-input '{
    "FederatedCatalog": {
        "Identifier": catalog-name-in-Snowflake",
        "ConnectionName": your-glue-connection-name"
    },
    "CreateTableDefaultPermissions": [],
    "CreateDatabaseDefaultPermissions": []
}'
```

Considerações ao integrar com Snowflake

- Quando você insere recursos (como bancos de dados e tabelas)Snowflake, o Lake Formation não revoga automaticamente as permissões concedidas nesse recurso federado. Para remover as permissões de acesso, você precisa revogar explicitamente as permissões que foram concedidas anteriormente no recurso federado usando o Lake Formation.
- Quando você monta o catálogo remoto com `CATALOG_CASING_FILTER='UPPERCASE_ONLY'` configuração, bancos de dados e tabelas com identificadores em maiúsculas são federados, mas objetos com identificadores em minúsculas não.
- Você pode reutilizar a mesma AWS Glue conexão para criar vários catálogos federados. A exclusão de um catálogo não excluirá a conexão associada. Para excluir uma conexão, use o `aws glue delete-connection` comando AWS CLI e certifique-se de que todos os catálogos associados sejam excluídos primeiro.
- Os namespaces aninhados no Polaris catálogo não são suportados. Ou seja, a federação de catálogos pode acessar Iceberg tabelas remotas que seguem uma notação `catalog.database.table` de três partes.

Federar para Catálogo do Unity do Databricks

AWS Glue Data Catalog federa para Databricks usar as credenciais do OAuth2 de um principal de serviço. Databricks Esse mecanismo de autenticação permite AWS Glue Data Catalog acessar os

metadados de vários objetos (como catálogos, bancos de dados e tabelas) emDatabricks Unity Catalog, com base nos privilégios associados ao principal do serviço. Para garantir o acesso aos objetos certos, é essencial conceder ao responsável pelo serviço as permissões necessárias Databricks para ler os metadados desses objetos.

Em seguida, a federação de catálogos permite a descoberta e a consulta de Iceberg tabelas em seu catálogo do Databricks Unity. Para ler tabelas delta, certifique-se de que Iceberg os metadados estejam disponíveis para o uso Uniform dessas tabelas. Siga o Databricks tutorial e a documentação para criar o principal de serviço e os privilégios associados em seu espaço Databricks de trabalho.

Pré-requisitos

Antes de criar um catálogo federado no Catálogo de Dados que seja controlado pelo Lake Formation, verifique se você tem as seguintes permissões:

Seu diretor do IAM (usuário ou função) deve ter as seguintes permissões:

- Permissões do Lake Formation —`lakeformation:RegisterResource`, `lakeformation:DescribeResource`
- AWS Glue permissões —`glue:CreateConnection`,`glue:CreateCatalog`,`glue:GetConnection`, `glue:PassConnection`
- Permissões do Secrets Manager —`secretsmanager:CreateSecret`, `secretsmanager:GetSecretValue`
- Permissões do IAM —`iam:CreateRole`,`iam:AttachRolePolicy`, `iam:PassRole`

Você deve ser administrador do data lake do Lake Formation ou ter `CREATE_CATALOG` permissão no catálogo de dados.

Criar catálogo federado

Usando o console

1. Faça login no console e abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.
2. Escolha a AWS região preferida no seletor de região.
3. No painel de navegação à esquerda, selecione Catálogos.

4. Escolha Criar catálogo para abrir o fluxo de trabalho Criar catálogo.
5. Na etapa Escolher fonte de dados, Databricks selecione entre as opções disponíveis.
6. Na etapa Definir detalhes do catálogo, você fornece três informações: detalhes do catálogo, detalhes da conexão e detalhes do registro.
7. No contêiner de detalhes do catálogo, forneça um nome exclusivo para seu catálogo AWS Glue federado e insira o nome do Databricks catálogo existente.
8. No contêiner de detalhes de conexões, você pode escolher entre uma conexão existente à qual você tem acesso ou fornecer configuração para criar um novo conector.
9. As novas configurações de conexão incluem:

    - Nome da conexão — Um nome exclusivo do objeto de AWS Glue conexão.
    - URL do espaço de trabalho — O URL do endpoint do seu espaço de trabalho existenteDatabricks.
    - Autenticação — Especifique a configuração de autenticação AWS Glue usada para se conectar ao servidor de catálogo remoto. AWS Glue suporta tanto o OAuth2 quanto a autenticação personalizada.
    - URL do token: especifique o URL do provedor de identidades do catálogo remoto.
    - ID do cliente do OAuth2 — especifique o ID do cliente da credencial do OAuth2 associada ao seu catálogo remoto.
    - Segredo — Armazene e use o segredo do cliente OAuth2 usando AWS Secrets Manager ou insira o valor do segredo na caixa de texto. Quando você insere o segredo manualmente no console, AWS Glue cria o segredo em seu nome.
    - Escopo do URL do token: especifique o escopo do OAuth para autenticação.
10. Crie uma função do IAM que AWS Glue os diretores de serviço do Lake Formation possam usar para acessar locais secretos e locais de tabelas Iceberg remotas no Amazon S3, respectivamente. Selecione a função do IAM no menu suspenso de registro. Consulte as etapas 2 e 3 na seção CLI a seguir para obter detalhes da política do IAM.
11. Selecione Testar conexão para testar se suas propriedades de conexão e o acesso à função do IAM estão configurados corretamente. A funcionalidade de teste de conexão não está disponível ao se conectar Databricks usando a Amazon VPC.
12. Selecione Avançar para revisar suas configurações.
13. Selecione Criar catálogo na página de revisão.

Usar a CLI

1. Crie um AWS Secrets Manager segredo

   O AWS Glue conector oferece suporte a dois tipos de autenticação: OAuth2 e Custom. Ao usar a opção OAuth2, use AWS Secrets Manager para armazenar o segredo do cliente do principal do Databricks serviço. Posteriormente, você usará esse segredo ao criar a AWS Glue Conexão. Para autenticação personalizada, use AWS Secrets Manager para armazenar e recuperar o token de acesso.

   No exemplo a seguir, substitua*<databricks-secret>*,*<client_secret>*, *<region>* por suas próprias informações.

   ```
   aws secretsmanager create-secret \
   --name <databricks-secret> \
   --description "Databricks secret" \
   --secret-string '{
   "USER_MANAGED_CLIENT_APPLICATION_CLIENT_SECRET": "<client_secret>"
   }' \
   --region <region>
   ```

   > **Note**
   >
   > `USER_MANAGED_CLIENT_APPLICATION_CLIENT_SECRET`é uma palavra-chave reservada AWS Glue usada para se referir a um valor secreto do cliente no segredo. Use a mesma palavra-chave ao criar o segredo no console do Lake Formation também.

2. Crie uma função do IAM que dê acesso ao objeto de AWS Glue conexão ao segredo criado na etapa anterior

   O objeto de AWS Glue conexão requer acesso ao AWS Secrets Manager segredo quando você o usa AWS Secrets Manager para armazenar, recuperar e atualizar seu token secreto do OAuth. O objeto de AWS Glue conexão também requer acesso para criar, descrever e usar as interfaces de rede da Amazon VPC quando você usa um endpoint da Amazon VPC para restringir a conectividade ao seu espaço de trabalho. Databricks

   Crie uma política do IAM e anexe-a a uma função do IAM. Adicione o principal de AWS Glue serviço à política de confiança.

No exemplo a seguir, substitua *<your-secrets-manager-ARN><your-vpc-id>*, e *<your-subnet-id1>* por suas próprias informações.

Example Política do IAM

```
{
    "Version": "2012-10-17",
    "Statement": [{
            "Effect": "Allow",
            "Action": [
                "secretsmanager:GetSecretValue",
                "secretsmanager:DescribeSecret",
                "secretsmanager:PutSecretValue"
            ],
            "Resource": [
                "<your-secrets-manager-ARN>"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "ec2:CreateNetworkInterface",
                "ec2:DeleteNetworkInterface",
                "ec2:DescribeNetworkInterfaces"
            ],
            "Resource": "*",
            "Condition": {
                "ArnEquals": {
                    "ec2:Vpc": "arn:aws:ec2:region:account-id:vpc/<your-vpc-id>",
                    "ec2:Subnet": ["arn:aws:ec2:region:account-id:subnet/<your-
subnet-id1>"]
                }
            }
        }
    ]
}
```

Example Política de confiança

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

```
        "Effect": "Allow",
        "Principal": {
            "Service": "glue.amazonaws.com"
        },
        "Action": "sts:AssumeRole"
    }]
}
```

3. Crie uma política do IAM que dê à Lake Formation acesso de leitura à localização do catálogo no Amazon S3

   Como proprietário de um catálogo federado no Catálogo de Dados, você usa o Lake Formation para conceder acesso a tabelas de baixa granularidade, acesso refinado em nível de coluna, nível de linha e nível de célula e acesso baseado em tags às suas equipes de dados. O Lake Formation usa uma função do IAM que lhe dá acesso às localizações subjacentes do Amazon S3 de suas tabelas remotas do Iceberg. Esse acesso permite que a Lake Formation forneça credenciais de acesso com escopo definido para mecanismos de análise que consultam tabelas remotas.

   Crie uma política do IAM e anexe-a a uma função do IAM. Adicione o diretor de serviço do Lake Formation à política de confiança da função do IAM.

   No exemplo a seguir, substitua *<your-s3-bucket-N>* e *<your-kms-key>* por suas próprias informações.

   Example Política do IAM

```
{
    "Version": "2012-10-17",
    "Statement": [{
            "Effect": "Allow",
            "Action": [
                "s3:GetObject"
            ],
            "Resource": [
                "arn:aws:s3:::<your-s3-bucket-1>/*",
                "arn:aws:s3:::<your-s3-bucket-2>/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
```

```
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::<your-s3-bucket-1>",
                "arn:aws:s3:::<your-s3-bucket-2>"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "kms:Decrypt",
                "kms:Encrypt"
            ],
            "Resource": [
                "<your-kms-key>"
            ]
        }
    ]
}
```

Example Política de confiança

```
{
    "Version": "2012-10-17",
    "Statement": [{
        "Sid": "",
        "Effect": "Allow",
        "Principal": {
            "Service": "lakeformation.amazonaws.com"
        },
        "Action": "sts:AssumeRole"
    }]
}
```

> **Note**
>
> Quando você usa o console do Lake Formation para criar um catálogo federado, o console usa uma única função do IAM com as duas políticas anexadas para concluir a configuração.

4. Criar um objeto de AWS Glue conexão

O Data Catalog suporta connectionTypeDATABRICKSICEBERGRESTCATALOG: para conectar o Data Catalog a. Databricks Esse AWS Glue conector é compatível com OAuth2 e métodos de autenticação personalizados.

O exemplo a seguir usa a configuração de autenticação OAuth2 para criar uma AWS Glue conexão. Substitua *highlighed sections* por suas informações.

```
aws glue create-connection \
    --connection-input '{
"Name": "<your-glue-connection-to-databricks-unity-account>",
"ConnectionType": "DATABRICKSICEBERGRESTCATALOG",
"ConnectionProperties": {
    "INSTANCE_URL": "<your-databricks-workspace-catalog-URL>",
    "ROLE_ARN": "<your-IAM-role-for-secrets-and-VPC-access>"
},
"AuthenticationConfiguration": {
    "AuthenticationType": "OAUTH2",
    "OAuth2Properties": {
        "OAuth2GrantType": "CLIENT_CREDENTIALS",
        "TokenUrl": "<your-internal-or-external-token-server-url>",
        "OAuth2ClientApplication": {
            "UserManagedClientApplicationClientId": "<your-client-id>"
        },
        "TokenUrlParametersMap": {
            "scope": "all-apis"
        }
    },
    "SecretArn": "arn:aws:secretsmanager:<aws-region>:<your-aws-account-
id>:secret:<databricks-secret>"
}
}'
```

5. Registre AWS Glue a conexão como um recurso do Lake Formation

   Usando o objeto de AWS Glue conexão (criado na Etapa 4) e a função do IAM (criada na Etapa 3), agora você pode registrar o objeto de AWS Glue conexão como um recurso gerenciado do Lake Formation.

   Substitua *<your-glue-connector-arn>* e *<your-IAM-role-ARN-having-LF-access>* por suas informações.

```
aws lakeformation register-resource \
    --resource-arn <your-glue-connector-arn> \
    --role-arn <your-IAM-role-ARN-having-LF-access> \
    --with-federation \
    --with-privileged-access
```

6. Crie um catálogo federado no Catálogo de Dados

   Depois de criar um objeto de AWS Glue conexão e registrá-lo no Lake Formation, você pode criar um catálogo federado no Catálogo de Dados:

   Forneça ao catálogo federado um nome exclusivo em*<your-federated-catalog-name>*, faça referência ao catálogo Databricks em *"<catalog-name-in-Databricks>* e insira o nome da conexão criada anteriormente em*"<your-glue-connection-name>*.

```
aws glue create-catalog \
    --name <your-federated-catalog-name> \
    --catalog-input '{
    "FederatedCatalog": {
        "Identifier": "<catalog-name-in-Databricks>",
        "ConnectionName": "<your-glue-connection-name>"
    },
    "CreateTableDefaultPermissions": [],
    "CreateDatabaseDefaultPermissions": []
}'
```

Considerações ao integrar com Databricks

- Quando você coloca recursos (como bancos de dados e tabelas) noDatabricks, o Lake Formation não revoga automaticamente as permissões concedidas nesse recurso federado. Para remover as permissões de acesso, você precisa revogar explicitamente as permissões que foram concedidas anteriormente no recurso federado usando o Lake Formation.
- Você pode consultar tabelas do Iceberg armazenadas no Amazon S3 usando essa integração. Ao usar qualquer outro formato de tabela ou armazenamento de objetos, você pode federar metadados em catálogos remotos AWS Glue e listar seus bancos de dados e tabelas, mas operações de consulta, como `SELECT ColumnFoo from TableBar` ocorrerão falhas durante a consulta com o erro “Falha ao ler a tabela”. Apache Iceberg O local de armazenamento de objetos não é suportado. '

- Você pode reutilizar a mesma AWS Glue conexão para criar vários catálogos federados. A exclusão de um catálogo não excluirá o objeto de conexão associado. Para excluir um objeto de conexão, use o `aws glue delete-connection` comando AWS CLI e certifique-se de que todos os catálogos associados sejam excluídos primeiro.

Trazendo dados do Amazon Redshift para o AWS Glue Data Catalog

Você pode gerenciar dados analíticos nos armazéns de dados do Amazon Redshift no (Catálogo de dados) e AWS Glue Data Catalog unificar os data lakes do Amazon S3 e os armazéns de dados do Amazon Redshift. O Amazon Redshift é um serviço de armazém de dados totalmente gerenciado em escala de petabytes na nuvem. AWS Um data warehouse do Amazon Redshift é um conjunto de recursos de computação chamados nós, que são organizados em um grupo chamado cluster. Cada cluster executa um mecanismo do Amazon Redshift e contém um ou mais bancos de dados.

No Amazon Redshift, é possível criar clusters provisionados e namespaces sem servidor do Amazon Redshift e registrá-los no Data Catalog. Ao fazer isso, você pode unificar dados no armazenamento gerenciado do Amazon Redshift (RMS) e nos buckets do Amazon S3 e acessar dados de mecanismos de analytics compatíveis com o Apache Iceberg.

Quando você registra namespaces e clusters, pode conceder acesso aos dados sem a necessidade de copiá-los ou movê-los. Para acessar mais informações sobre o registro de clusters e namespaces no Amazon Redshift, consulte Registrar clusters e namespaces do Amazon Redshift no AWS Glue Data Catalog.

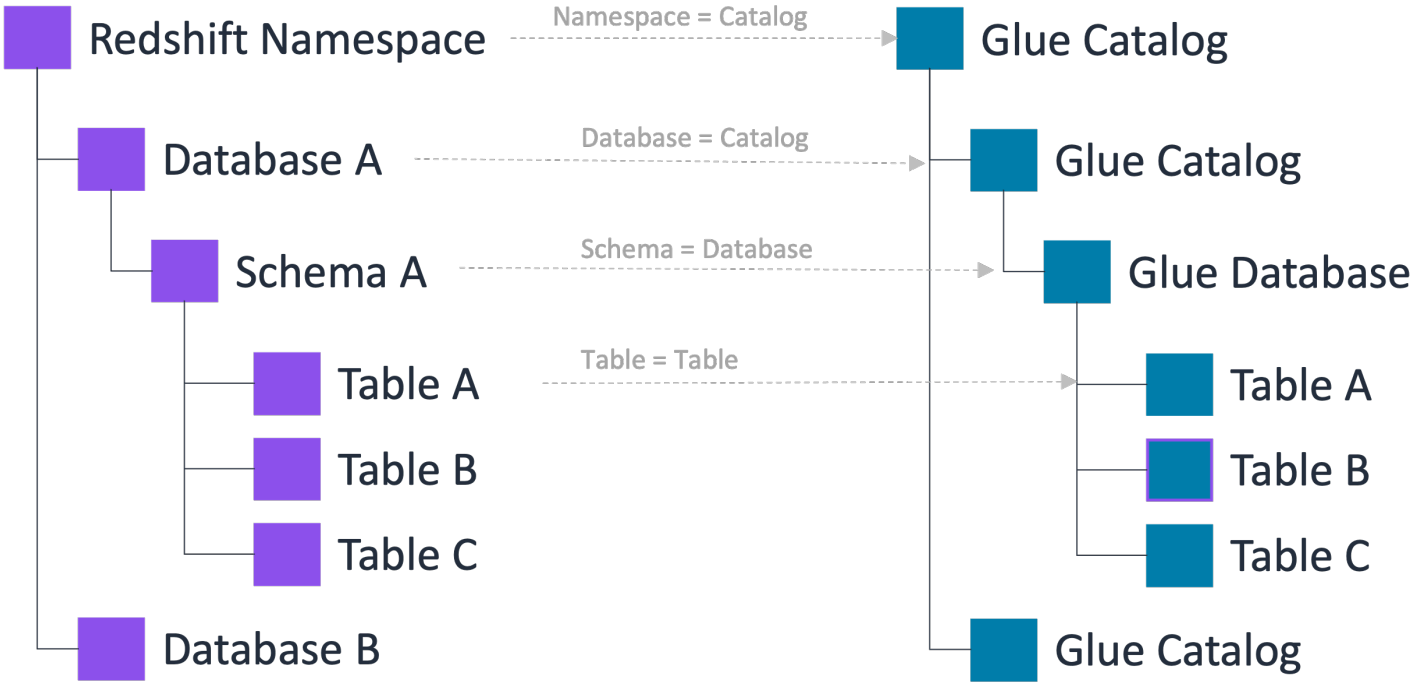
No Amazon Redshift, é possível realizar o compartilhamento de dados por meio de unidades de compartilhamento de dados ou registrar namespaces e clusters com o Data Catalog. Com as unidades de compartilhamento de dados, que operam no nível de objeto de banco de dados individual, você precisa habilitar o compartilhamento para cada tabela ou visualização. Por outro lado, a publicação de namespace funciona em nível de cluster ou namespace. Quando você registra um cluster ou namespace no Data Catalog, todos os bancos de dados e tabelas dentro dele são compartilhados automaticamente, sem que você precise configurar o compartilhamento para objetos individuais.

No Data Catalog, você pode criar um catálogo federado para cada namespace ou cluster. Um catálogo é chamado de catálogo federado quando aponta para uma entidade fora do Data Catalog. As tabelas e visualizações no namespace do Amazon Redshift são listadas como tabelas individuais

no Data Catalog. É possível compartilhar bancos de dados e tabelas no catálogo federado com entidades principais do IAM e usuários SAML selecionados na mesma conta ou em outra conta com Lake Formation. Você também pode incluir expressões de filtro de linha e coluna para restringir o acesso a determinados dados. Para obter mais informações, consulte Filtragem de dados e segurança por célula no Lake Formation.

O Data Catalog comporta uma hierarquia de metadados de três níveis que inclui catálogos, bancos de dados e tabelas (e visualizações). Quando você registra um namespace no Data Catalog, a hierarquia de dados do Amazon Redshift é associada à hierarquia de três níveis do Data Catalog da seguinte forma:

- O namespace Amazon Redshift se torna um catálogo de vários níveis no Data Catalog.
- O banco de dados associado do Amazon Redshift é registrado como um catálogo no Data Catalog.
- O esquema do Amazon Redshift se torna um banco de dados no Data Catalog.
- A tabela do Amazon Redshift se torna uma tabela no Data Catalog.



Com essa hierarquia de metadados de três níveis, você pode acessar as tabelas do Amazon Redshift usando a notação em três partes - "catalog1/catalog2.database.table” no catálogo de dados. Além disso, as equipes de dados podem manter a mesma organização que o Amazon Redshift usa para organizar tabelas na conta do Data Catalog.

No Lake Formation, você pode gerenciar com segurança os dados do Amazon Redshift usando controle de acesso refinado para os recursos do Data Catalog. Com essa integração, é possível gerenciar, proteger e consultar dados analíticos de um único catálogo com um mecanismo de controle de acesso comum.

Para conhecer as limitações, consulte Limitações para trazer dados do armazém de dados do Amazon Redshift para o AWS Glue Data Catalog.

Tópicos

- Benefícios principais
- Perfis e responsabilidades
- Pré-requisitos para gerenciar namespaces do Amazon Redshift no AWS Glue Data Catalog
- Criar catálogos federados do Amazon Redshift
- Visualizar objetos do catálogo
- Atualizar um catálogo federado
- Acessar um catálogo federado compartilhado
- Excluir um catálogo federado
- Consultar catálogos federados
- Recursos adicionais do

Benefícios principais

O registro de clusters e namespaces do Amazon Redshift com eles e AWS Glue Data Catalog a unificação de dados nos data lakes do Amazon S3 e nos armazéns de dados do Amazon Redshift oferecem os seguintes benefícios:

- Experiência de consulta uniforme: consulte seus dados e dados gerenciados pelo Amazon Redshift nos buckets do Amazon S3 usando qualquer mecanismo de consulta compatível com o Apache Iceberg, como o Amazon EMR Sem Servidor e o Amazon Athena, sem precisar mover ou copiar dados.
- Acesso consistente aos dados em todos os serviços — Você não precisa atualizar os nomes do banco de dados e das tabelas em seus pipelines de dados ao acessar as mesmas fontes de dados federadas de diferentes serviços de AWS análise, pois as fontes de dados são registradas no Catálogo de Dados.

- Fine-grained controle de acesso — Você pode aplicar as permissões do Lake Formation para gerenciar o acesso às fontes de dados federadas usando permissões de controle de acesso refinadas.

Perfis e responsabilidades

Perfil	Responsabilidade
Administrador de cluster de produtor do Amazon Redshift	Registra o cluster ou o namespace no Data Catalog.
Administrador do data lake do Lake Formation	Aceita o convite do cluster ou do namespace , cria catálogos federados e concede acesso aos catálogos federados a outras entidades principais.
Administrador somente leitura do Lake Formation	Descobre o catálogo federado e consulta as tabelas do Amazon Redshift no catálogo federado.
Perfil de transferência de dados	O Amazon Redshift assume, em seu nome, a transferência de dados para e do bucket do Amazon S3.

Veja a seguir as etapas de alto nível para conceder acesso aos usuários a um namespace do Amazon Redshift:

1. No Amazon Redshift, o administrador do cluster produtor registra um cluster ou namespace no Data Catalog.
2. O administrador do data lake aceita o convite do namespace do administrador do cluster produtor do Amazon Redshift e cria um catálogo federado no Data Catalog.

   Após concluir essa etapa, você pode gerenciar o catálogo de namespaces do Amazon Redshift dentro do Data Catalog.

3. Conceda permissões a usuários em catálogos, bancos de dados e tabelas. Você pode compartilhar todo o catálogo de namespaces ou um subconjunto de tabelas com os usuários na mesma conta ou em outra conta.

Pré-requisitos para gerenciar namespaces do Amazon Redshift no AWS Glue Data Catalog

1. Crie um administrador de data lake — crie uma função do IAM autorizada a aceitar o convite do namespace, crie os AWS Glue Data Catalog objetos (catálogos, bancos de dados tables/views) e conceda permissões do Lake Formation a outros usuários.

   Para obter instruções passo a passo sobre como criar um administrador de data lake, consulte Crie um administrador de data lake.

2. Atualize as permissões do administrador do data lake.

   Além das permissões, o administrador do data lake precisa das seguintes permissões para aceitar um convite de namespace do Amazon Redshift no Lake Formation, criar ou atualizar os recursos do Data Catalog e habilitar o acesso ao data lake:

   JSON

```
{
"Version":"2012-10-17",
"Id": "glue-enable-datalake-access",
"Statement": [{
        "Effect": "Allow",
        "Action": [
            "redshift:AssociateDataShareConsumer",
            "redshift:DescribeDataSharesForConsumer",
            "redshift:DescribeDataShares",
            "redshift-serverless:CreateNamespace",
            "redshift-serverless:CreateWorkgroup",
            "redshift-serverless:DeleteNamespace",
            "redshift-serverless:DeleteWorkgroup",
            "ec2:DescribeAccountAttributes",
            "ec2:DescribeSubnets",
            "ec2:DescribeSecurityGroups",
            "ec2:DescribeAvailabilityZones",
            "s3:createBucket",
```

```
                "s3:deleteBucket",
                "s3:putBucketPolicy",
                "s3:putEncryptionConfiguration",
                "s3:putLifecycleConfiguration",
                "s3:putBucketVersioning",
                "iam:CreateRole"
            ],
            "Resource": "*"
        },
        {
            "Action": [
                "iam:PassRole"
            ],
            "Effect": "Allow",
            "Resource": "arn:aws:iam::*:role/data transfer role name",
            "Condition": {
                "StringLike": {
                    "iam:PassedToService": [
                        "glue.amazonaws.com"
                    ]
                }
            }
        }
    ]
}
```

3. Se o perfil do IAM usado para criar catálogos federados não for de administrador de data lake, você precisará conceder a permissão `Create catalog` ao perfil.

   Como criar criadores de catálogos

   a. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

   b. Escolha Perfis e tarefas administrativas em Administração.

   c. Selecione Conceder.

   d. Na tela Conceder permissões, selecione um usuário ou um perfil do IAM.

   e. Selecione a permissão Criar catálogo.

   f. Você também pode conceder a permissão Criar catálogo. A permissão para concessão permite que o criador de catálogos conceda a permissão `Create catalog` a outras entidades principais.

   g. Selecione Conceder.

AWS CLI exemplo de concessão de permissões para criar um catálogo federado.

```
aws lakeformation grant-permissions \
--cli-input-json \
'{
    "Principal": {
     "DataLakePrincipalIdentifier":"arn:aws:iam::123456789012:role/Admin"
    },
    "Resource": {
        "Catalog": {
        }
    },
    "Permissions": [
        "CREATE_CATALOG",
        "DESCRIBE"
    ]
}'
```

4. Crie um perfil de administrador somente leitura para descobrir os catálogos federados do Amazon Redshift no Data Catalog do Editor de Consultas V2 do Amazon Redshift.

   Para consultar as tabelas do Amazon Redshift no catálogo federado do Editor de Consultas V2 do Amazon Redshift, a política de perfil de administrador somente leitura deve conter o ARN do perfil vinculado a serviços `AWSServiceRoleForRedshift` do Amazon Redshift.

```
 aws lakeformation put-data-lake-settings
        --region us-east-1 \
        --data-lake-settings \
 '{
   "DataLakeAdmins": [{"DataLakePrincipalIdentifier":"arn:aws:iam::123456789012:role/
Admin"}],
   "ReadOnlyAdmins":[{"DataLakePrincipalIdentifier":"arn:aws:iam::123456789012:role/
aws-service-role/redshift.amazonaws.com/AWSServiceRoleForRedshift"}],
   "CreateDatabaseDefaultPermissions":[],
   "CreateTableDefaultPermissions":[],
   "Parameters":{"CROSS_ACCOUNT_VERSION":"4","SET_CONTEXT":"TRUE"}
  }'
```

5. Crie um perfil de transferência de dados que o Amazon Redshift possa assumir em seu nome para transferir dados de e para o bucket do Amazon S3.

Ao habilitar o acesso ao data lake para mecanismos de consulta compatíveis com o Apache Iceberg, como Athena, Amazon EMR no Amazon EC2, para acessar os recursos do Amazon Redshift no Data Catalog, você precisa criar um perfil do IAM com as permissões necessárias para realizar a transferência de dados de e para o bucket do Amazon S3.

JSON

```
{
    "Version":"2012-10-17",
    "Id": "glue-enable-datalake-access",
    "Statement": [{
      "Sid": "DataTransferRolePolicy",
        "Effect": "Allow",
        "Action": [ "glue:GetCatalog",
                    "glue:GetDatabase",
                    "kms:GenerateDataKey",
                    "kms:Decrypt"],
        "Resource": "*"
    }
    ]
}
```

6. Adicione a seguinte política de confiança à função de transferência de dados AWS Glue e aos serviços do Amazon Redshift para assumir a função de transferir dados de e para o bucket do Amazon S3.

   JSON

```
      {
    "Version":"2012-10-17",
    "Statement": [{
        "Effect": "Allow",
        "Principal": {
        "Service": [
        "redshift.amazonaws.com",
        "glue.amazonaws.com"
        ]
    },
        "Action": "sts:AssumeRole"
    }]
```

```
}
```

7. Adicione a seguinte política de chaves à AWS KMS chave se você estiver usando uma chave gerenciada pelo cliente para criptografar os dados no Amazon cluster/namespace Redshift. Substitua o número da conta por um número de AWS conta válido e especifique o nome da função de transferência de dados. Por padrão, os dados no cluster do Amazon Redshift são criptografados com uma chave do KMS. O Lake Formation oferece uma opção para criar sua chave do KMS personalizada para criptografia. Se você estiver usando uma chave gerenciada pelo cliente, deverá adicionar políticas de chave específicas à chave.

   Para obter mais informações sobre como gerenciar as permissões de uma chave gerenciada pelo cliente, consulte Chaves gerenciadas pelo cliente.

   JSON

```
{
    "Version":"2012-10-17",
    "Id": "auto-redshift-3",
    "Statement": [{
            "Sid": "RedshiftAllowAccessPolicy",
            "Effect": "Allow",
            "Principal": {
                "AWS": "*"
            },
            "Action": [
                "kms:Encrypt",
                "kms:Decrypt",
                "kms:ReEncrypt*",
                "kms:GenerateDataKey*",
                "kms:CreateGrant",
                "kms:DescribeKey"
            ],
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "kms:CallerAccount": "111122223333",
                    "kms:ViaService": "redshift.us-east-1.amazonaws.com"
                }
            }
        },
        {
            "Sid": "RedshiftServerlessAllowAccessPolicy",
```

```
            "Effect": "Allow",
            "Principal": {
                "AWS": "*"
            },
            "Action": [
                "kms:Encrypt",
                "kms:Decrypt",
                "kms:ReEncrypt*",
                "kms:GenerateDataKey*",
                "kms:CreateGrant",
                "kms:DescribeKey"
            ],
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "kms:CallerAccount": "111122223333",
                    "kms:ViaService": "redshift-serverless.us-
east-1.amazonaws.com"
                }
            }
        },
        {
            "Sid": "DirectMetadataAccess",
            "Effect": "Allow",
            "Principal": {
            "AWS": "arn:aws:iam::111122223333:root"
            },
            "Action": [
                "kms:Describe*",
                "kms:Get*",
                "kms:List*",
                "kms:RevokeGrant"
            ],
            "Resource": "*"
        },
        {
            "Sid": "GenerateDataKeyDecryptDataTransferRole",
            "Effect": "Allow",
            "Principal": {
            "AWS": "arn:aws:iam::111122223333:role/data-transfer-role-name"
            },
            "Action": [
                "kms:GenerateDataKey",
                "kms:Decrypt"
```

```
            ],
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                "kms:ViaService": "s3.us-east-1.amazonaws.com"
                }
            }
        }
    ]
}
```

Criar catálogos federados do Amazon Redshift

Este tópico descreve as etapas que você precisa seguir para aceitar um convite de cluster ou namespace, criar um catálogo federado de vários níveis e conceder permissões a outras entidades principais. Você pode concluir essas tarefas usando o console Lake Formation, o AWS Command Line Interface (AWS CLI) ou APIs/SDKs. Os exemplos neste tópico mostram o produtor cluster/namespace, o catálogo de dados e o consumidor de dados na mesma conta.

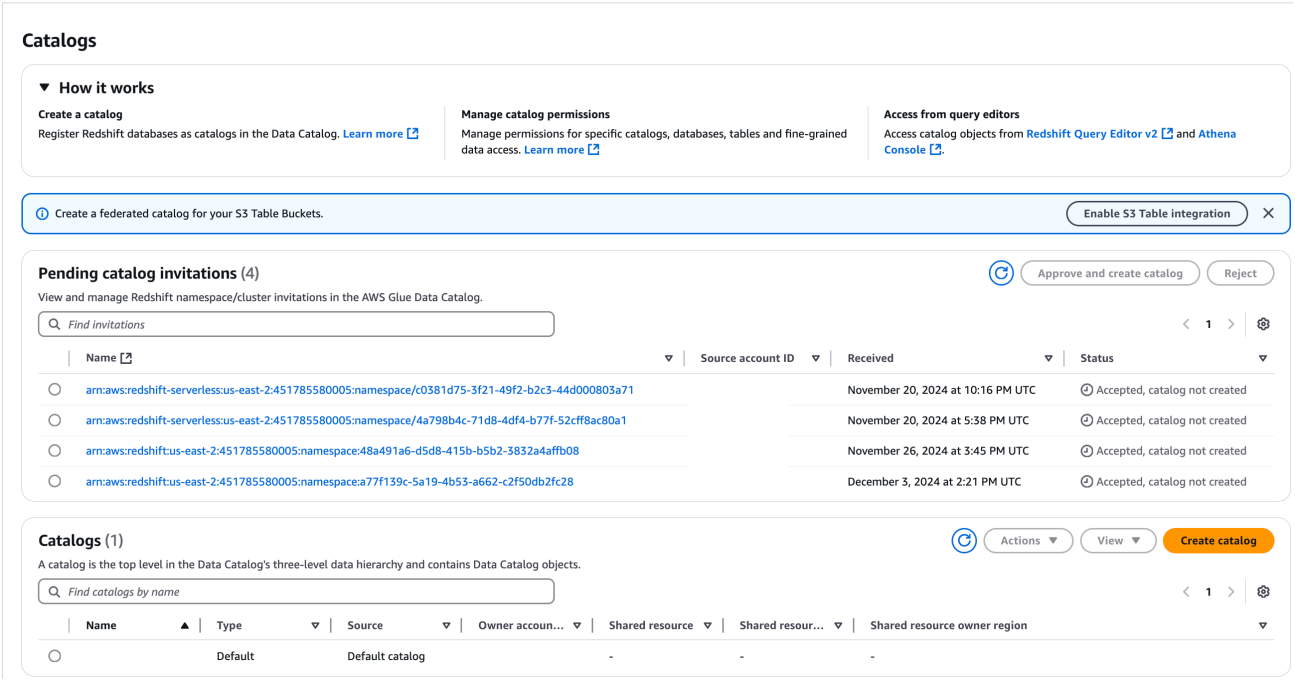
Para saber mais sobre os recursos de entre contas do Lake Formation, consulte Cross-account compartilhamento de dados em Lake Formation.

Como gerenciar um namespace do Amazon Redshift no Data Catalog

1. Analise um convite de namespace e aceite-o.

   Console

   1. Faça login no console do Lake Formation como administrador do data lake em https://console.aws.amazon.com/lakeformation/. Navegue até a página Catálogos em Data Catalog.
   2. Analise o convite do namespace que você tem autorização para acessar. A coluna Status indica seu status atual de participação no namespace. O status Não aceito indica que você foi adicionado ao namespace, mas não o aceitou ainda ou rejeitou o convite.



3. Para responder ao convite de um namespace ou cluster, selecione o nome do convite e escolha Revisar convite. Em Aceitar ou rejeitar convite, examine os detalhes do convite. Selecione Aceitar para aceitar o convite ou Rejeitar para recusar o convite. Se você rejeitar o convite, não terá acesso ao namespace.

AWS CLI

Os exemplos a seguir mostram como exibir, aceitar e registrar o convite. Substitua o Conta da AWS ID por um Conta da AWS ID válido. Substitua o `data-share-arn` pelo nome do recurso da Amazon (ARN) real que faz referência ao namespace.

1. Visualize um convite pendente.

```
aws redshift describe-data-shares \
 --data-share-arn 'arn:aws:redshift:us-
east-1:123456789012:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
ds_internal_namespace' \
```

2. Aceite um convite.

```
 aws redshift associate-data-share-consumer \
 --data-share-arn 'arn:aws:redshift:us-
east-1:123456789012:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
ds_internal_namespace' \
```

```
 --consumer-arn 'arn:aws:glue:us-east-1:123456789012:catalog'
```

3. Registre o cluster ou o namespace na conta do Lake Formation. Use a operação RegisterResourceda API para registrar o compartilhamento de dados no Lake Formation. `DataShareArn`é o parâmetro de entrada para`ResourceArn`.

> **Note**
>
> Este é uma etapa obrigatória.

```
aws lakeformation register-resource \
 --resource-arn 'arn:aws:redshift:us-
east-1:123456789012:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
ds_internal_namespace'
```

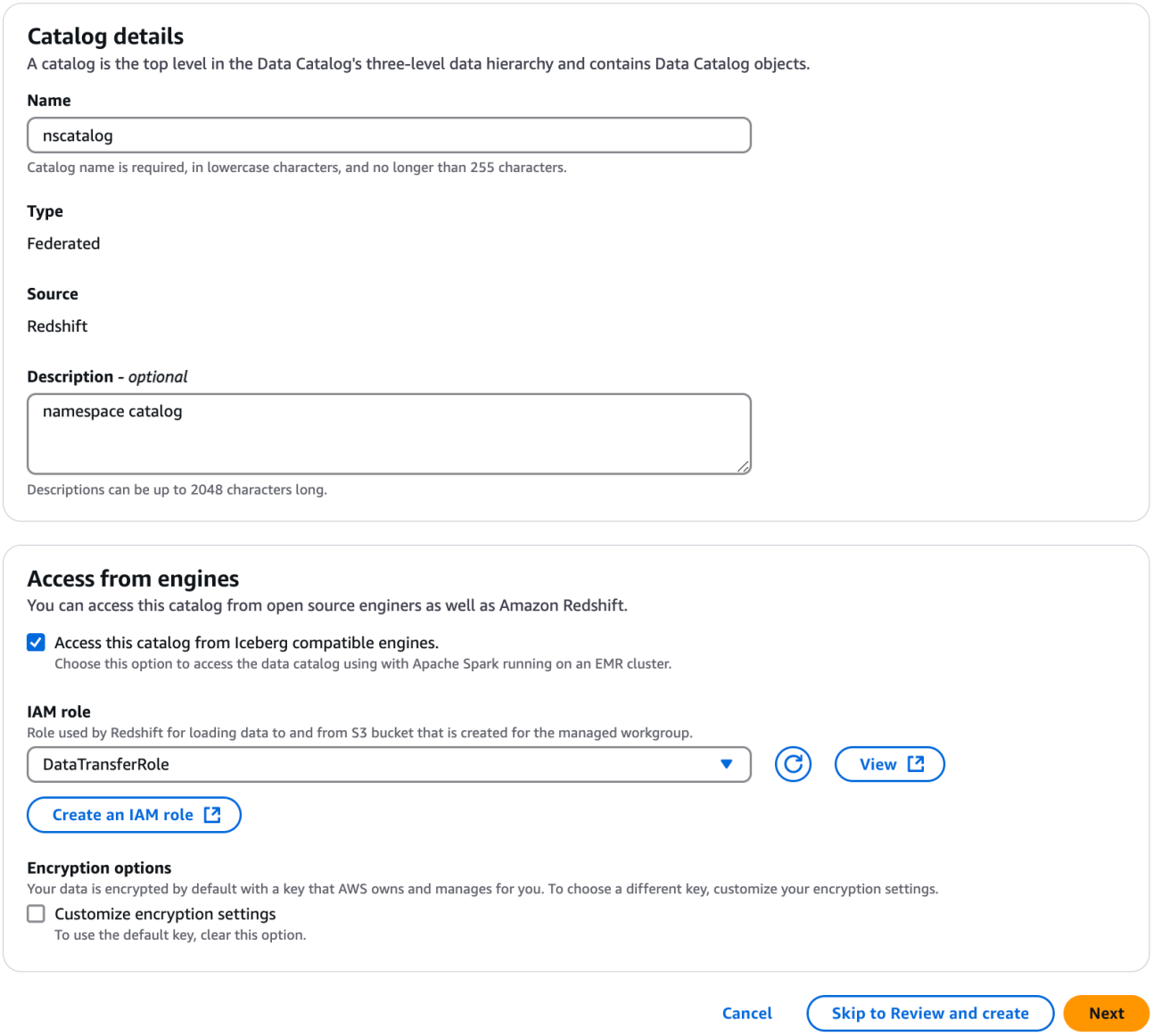
2. Crie um catálogo federado.

Depois de aceitar um convite, você precisa criar um catálogo federado no Data Catalog que associe os objetos no namespace do Amazon Redshift ao Data Catalog. Você deve ser administrador de data lake ou um usuário ou perfil que tenha as permissões necessárias para criar um catálogo.

Console

1. Depois de aceitar o convite do namespace, a página Definir detalhes do catálogo é exibida.
2. Na página Definir detalhes do catálogo, insira um nome exclusivo para o catálogo. Use letras minúsculas para nomes de catálogos. Os nomes dos catálogos devem ter, no máximo, 255 caracteres. Você usa esse identificador para mapear o namespace internamente na hierarquia de metadados (catalogid.db). Name.schema.table
3. Digite uma descrição para o catálogo. A descrição deve ter 2.048 caracteres ou menos.
4. Depois, marque a caixa de seleção Acessar esse catálogo por meio de mecanismos compatíveis com o Iceberg para permitir o acesso aos recursos do Amazon Redshift usando mecanismos analíticos compatíveis com o Apache Iceberg, como o Athena e o Apache Spark no Amazon EMR.

   Não é necessário habilitar o acesso ao data lake para acessar os catálogos federados usando o Amazon Redshift.



5. Para permitir que esses mecanismos de consulta leiam e gravem nos namespaces do Amazon Redshift, AWS Glue cria um cluster gerenciado do Amazon Redshift com os recursos computacionais e de armazenamento necessários para realizar operações de leitura e gravação sem afetar as cargas de trabalho do armazém de dados do Amazon Redshift.

   Também é preciso fornecer um perfil do IAM com as permissões necessárias para transferir dados de e para o bucket do Amazon S3.

6. Por padrão, os dados no cluster do Amazon Redshift são criptografados usando uma chave AWS gerenciada. O Lake Formation oferece uma opção para criar sua chave do KMS personalizada para criptografia. Se você estiver usando uma chave gerenciada pelo cliente, deverá adicionar políticas de chave específicas à chave.

Escolha a opção Personalizar configurações de criptografia se você estiver usando uma chave gerenciada pelo cliente para criptografar os dados no Amazon cluster/namespace Redshift. Para usar uma chave personalizada, você deve adicionar outra política de chave gerenciada personalizada à sua chave do KMS. Para obter mais informações, consulte Pré-requisitos para gerenciar namespaces do Amazon Redshift no AWS Glue Data Catalog.

AWS CLI

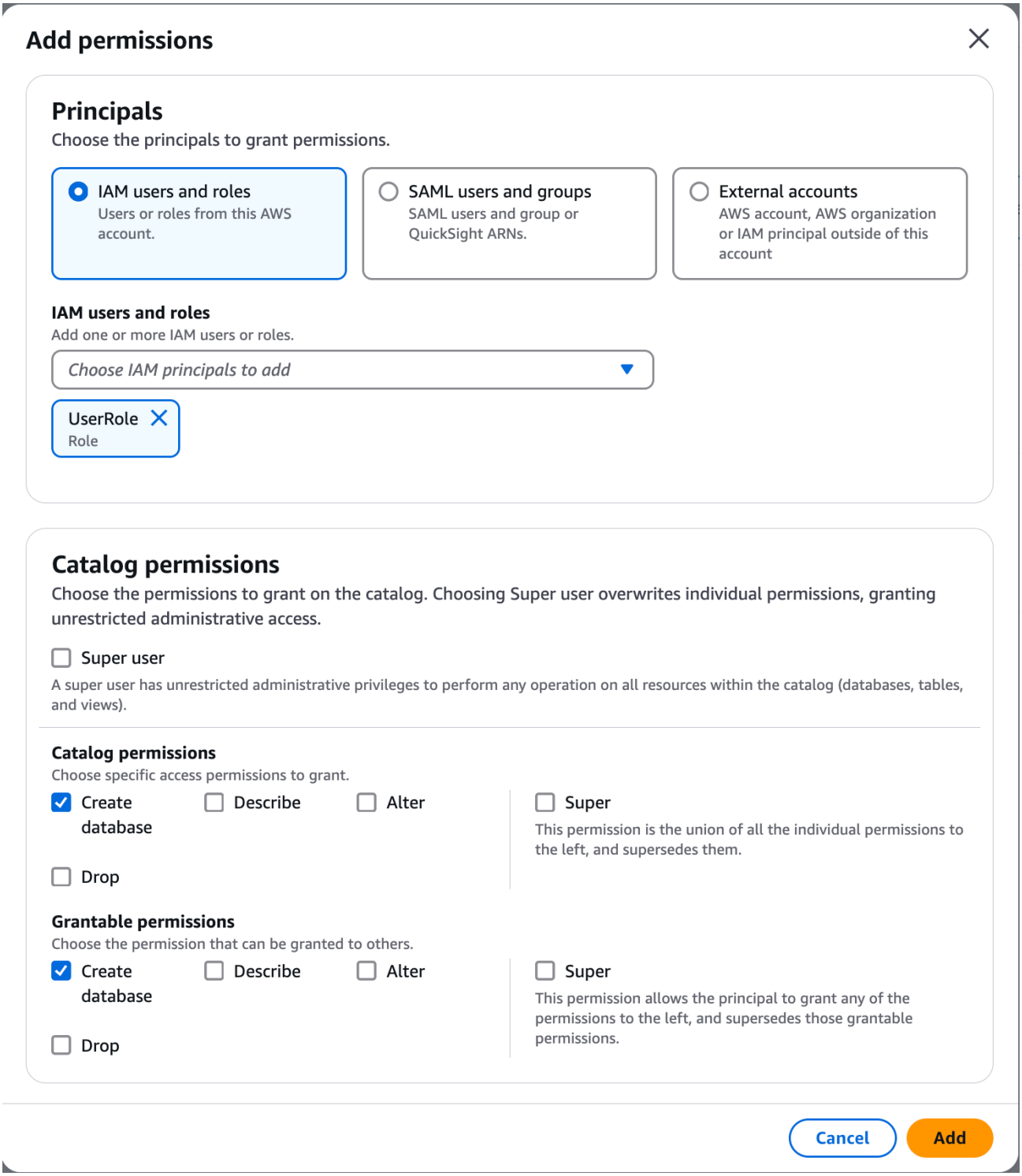
Use o código de exemplo a seguir para criar um catálogo com os dados do Amazon Redshift publicados no Data Catalog usando a AWS CLI.

```
aws glue create-catalog
--cli-input-json \
'{
    "Name": "nscatalog",
    "CatalogInput": {
        "Description": "Redshift federated catalog",
        "CreateDatabaseDefaultPermissions" : [],
        "CreateTableDefaultPermissions": [],
        "FederatedCatalog": {
            "Identifier": "arn:aws:redshift:us-
east-1:123456789012:datashare:11524d7f-f56d-45fe-83f7-d7bb0a4d6d71/
ds_internal_namespace",
            "ConnectionName": "aws:redshift"
        },
        "CatalogProperties": {
          "DataLakeAccessProperties" : {
            "DataLakeAccess" : true,
            "DataTransferRole" :
 "arn:aws:iam::123456789012:role/DataTransferRole"
         }
       }
    }
}'
```

3. Conceda permissões aos usuários em sua conta ou em contas externas.

Console de gerenciamento da AWS

1. Selecione Próximo para conceder permissões a outros usuários nos catálogos, banco de dados e tabelas compartilhados.
2. Na tela Adicionar permissões, escolha as entidades principais e os tipos de permissões a serem concedidas.



a. Na seção Entidades principais, escolha um tipo de entidade principal e, em seguida, especifique as entidades principais para conceder permissões.

   - Usuários e perfis do IAM: escolha um ou mais usuários ou perfis na lista de usuários e perfis do IAM.

- Usuários e grupos do SAML — Para SAML e Amazon Quick usuários e grupos, insira um ou mais Amazon Resource Names (ARNs) para usuários ou grupos federados por meio do SAML, ou ARNs para usuários ou grupos do Amazon Quick. Pressione Enter após cada ARN.

  Para obter informações sobre como construir os ARNs, consulte os comandos AWS CLI grant and revoke AWS CLI .
- Contas externas — Para AWS, AWS organização ou diretor do IAM, insira um ou mais IDs de AWS conta, IDs de organização, IDs de unidade organizacional ou ARN válidos para o usuário ou função do IAM. Pressione Enter após cada ID. O ID da organização consiste em “o-” seguido por 10 a 32 letras minúsculas ou dígitos. Uma ID de unidade organizacional começa com “ou-” seguida de 4 a 32 letras minúsculas ou dígitos (o ID da raiz que contém a OU). Essa sequência é seguida por um segundo travessão "-" e 8 a 32 letras minúsculas ou dígitos adicionais.

b. Na seção Permissões, selecione permissões e permissões concedidas.

   Em Permissões do catálogo, selecione uma ou mais permissões a serem concedidas. Em Permissões concedidas, selecione as permissões que o beneficiário do subsídio pode conceder a outros diretores em sua conta. AWS Essa opção não é compatível quando você está concedendo permissões a uma entidade principal do IAM a partir de uma conta externa.

   Escolha Superusuário para conceder ao usuário permissões irrestritas aos recursos (bancos de dados, tabelas, visualizações) dentro do catálogo.

3. Escolha Adicionar.

AWS CLI

Use os exemplos a seguir para conceder permissões de catálogo, banco de dados e tabelas usando a AWS CLI:

- O exemplo a seguir mostra como conceder permissões no catálogo federado.

```
aws lakeformation grant-permissions
 --cli-input-cli-json \
   '{
         "Principal": {
```

```
              "DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:role/
non-admin"
          },
          "Resource": {
               "Catalog": {
                     "Id": "123456789012:nscatalog"
                }
            },
            "Permissions": [
                   "DESCRIBE","CREATE_CATALOG"
             ],
            "PermissionsWithGrantOption": [
             ]
    }'
```

- Use o exemplo da seguir para conceder permissões em um banco de dados.

```
aws lakeformation grant-permissions \
  --cli-input-json \
          '{
              "Principal": {

 "DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:role/non-admin"
              },
              "Resource": {
                  "Database": {
                      "CatalogId": "123456789012:nscatalog/dev",
                      "Name": "public"
                  }
              },
              "Permissions": [
                  "ALL"
              ]
          }'
```

- O exemplo a seguir mostra como conceder permissões em uma tabela no banco de dados do Amazon Redshift.

```
aws lakeformation grant-permissions \
  --cli-input-json \
        '{
            "Principal": {
```

```
"DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:role/non-admin"
            },
            "Resource": {
                "Table": {
                    "CatalogId": "123456789012:nscatalog2/dev",
                    "DatabaseName": "public",
                    "TableWildcard" : {}
                }
            },
            "Permissions": [
                "ALL"
            ]
        }'
```

4. Escolha Próximo para revisar os detalhes do catálogo e criar um catálogo federado. O catálogo federado recém-criado e os objetos do catálogo aparecem na página Catálogos.

   Um catálogo federado do Amazon Redshift é referido com `catalogID = 123456789012:Redshift-federated catalog id`.

Visualizar objetos do catálogo

Depois de criar o catálogo federado, você pode visualizar os objetos nele usando o console do Lake Formation ou a AWS CLI.

Console de gerenciamento da AWS

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.
2. Escolha Catálogos no Data Catalog.
3. Escolha um catálogo federado na lista na página Catálogos.
4. A página de resumo do catálogo mostra os objetos do catálogo (bancos de dados e tabelas) nos quais você tem permissões. A guia Permissões mostra as entidades principais do IAM que receberam permissões nesses objetos.

AWS CLI

- O AWS CLI exemplo a seguir mostra como solicitar o catálogo de nível superior.

```
aws glue get-catalog \
--catalog-id 123456789012:nscatalog
```

Resposta

```
{
    "Catalog": {
        "CatalogId": "123456789012:nscatalog",
        "Name": "nscatalog",
        "ResourceArn": "arn:aws:glue:us-east-1:123456789012:catalog/nscatalog",
        "Description": "Redshift published Catalog",
        "CreateTime": "2024-09-05T14:49:16-07:00",
        "FederatedCatalog": {
            "Identifier": "arn:aws:redshift:us-
east-1:123456789012:datashare:b1234589-e823-4a14-ad8e-077085540a50/
ds_internal_namespace",
            "ConnectionName": "aws:redshift"
        },
        "CatalogProperties": {
            "DataLakeAccessProperties": {
                "DataLakeAccess": true,
                "DataTransferRole": "arn:aws:iam::123456789012:role/
DataTransferRole",
                "KmsKey": "AWS_OWNED_KMS_KEY",
                "ManagedWorkgroupName": "123456789012:nscatalog",
                "ManagedWorkgroupStatus": "AVAILABLE",
                "RedshiftDatabaseName": "dev"
            }
        },
        "CatalogIdentifier": "e2309c2c2fb048f1a3069dfdc1c7883e",
        "CreateTableDefaultPermissions": [],
        "CreateDatabaseDefaultPermissions": []
    }
}
```

- O exemplo a seguir mostra como solicitar todos os catálogos na conta.

```
aws glue get-catalogs \
  --recursive
```

- O exemplo de solicitação a seguir mostra como acessar um catálogo em nível de banco de dados do Amazon Redshift.

```
aws glue get-catlog \
 --catalog-id 123456789012:namespace catalog name/redshift database name
```

- O exemplo de solicitação a seguir mostra como acessar os bancos de dados no catálogo em nível de banco de dados do Amazon Redshift.

```
aws glue get-databases \
--catalog-id 123456789012:namespace catalog name/redshift database name
```

- O exemplo de solicitação a seguir mostra como acessar uma tabela do Amazon Redshift no catálogo.

```
aws glue get-table \
  --catalog-id 123456789012:parent catalog name/redshift database \
  --database-name redshift schema name \
  --name table name
```

- O exemplo de código a seguir mostra como acessar todas as tabelas do banco de dados do Amazon Redshift.

```
aws glue get-tables \
 --catalog-id 123456789012:namespace catalog name/redshift database name \
 --database-name RS schema name
```

Atualizar um catálogo federado

Você pode atualizar um catálogo federado do Amazon Redshift no catálogo de dados usando o console do Lake Formation AWS CLI ou a operação da UpdateCatalogAPI.

Console de gerenciamento da AWS

Siga estas etapas para atualizar seu catálogo federado usando o console do Lake Formation.

1. Faça login no Console de gerenciamento da AWS, e abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

2. No painel de navegação esquerdo, escolha Catálogos, em Data Catalog.
3. Na página Catálogos, escolha o catálogo federado do Amazon Redshift que você deseja atualizar.
4. Em Ações, selecione Editar.
5. Na tela Definir detalhes do catálogo, na seção Acesso a partir de mecanismos, escolha Acessar este catálogo por meio de mecanismos compatíveis com o Iceberg. Marcar essa opção habilitará o acesso ao data lake para mecanismos de consulta compatíveis com o Apache Iceberg.
6. Depois, crie um perfil do IAM ou escolha um perfil existente do IAM que tenha a política que concede permissões para realizar a transferência de dados de e para o bucket do Amazon S3.

   Para acessar mais informações sobre permissões, consulte Pré-requisitos para gerenciar namespaces do Amazon Redshift no AWS Glue Data Catalog.
7. Por padrão, os dados no cluster do Amazon Redshift são criptografados com uma Chave gerenciada pela AWS. Se você optar por criptografar dados usando uma chave gerenciada pelo cliente, crie uma chave do KMS ou escolha uma chave existente que tenha as permissões definidas na seção Pré-requisitos para gerenciar namespaces do Amazon Redshift no AWS Glue Data Catalog.
8. Escolha Salvar.

   Após a conclusão bem-sucedida, a página Detalhes do catálogo mostra o nome do grupo de trabalho gerenciado com o status “Êxito”.

AWS CLI

Veja a seguir um exemplo da entrada da CLI `update-catalog` com o acesso ao data lake desabilitado ao definir o valor do parâmetro `DataLakeAacess` como `false`.

```
aws glue update-catalog  --cli-input-json \
'{
    "Name": "nscatalog",
    "CatalogInput": {
        "Description": "Redshift published catalog",
        "CreateDatabaseDefaultPermissions" : [],
        "CreateTableDefaultPermissions": [],
        "FederatedCatalog": {
```

```
            "Identifier": "arn:aws:redshift:us-
east-1:123456789012:datashare:11524d7f-f56d-45fe-83f7-d7bb0a4d6d71/
ds_internal_namespace",
            "ConnectionName": "aws:redshift"
        },
        "CatalogProperties": {
          "DataLakeAccessProperties" : {
            "DataLakeAccess" : false
        }
       }
    }
}'
```

Acessar um catálogo federado compartilhado

AWS Lake Formation os recursos de várias contas permitem que os usuários compartilhem com segurança lagos de dados distribuídos entre vários Conta da AWS s, AWS organizações ou diretamente com os diretores do IAM em outra conta, fornecendo acesso refinado aos metadados e aos dados subjacentes.

Lake Formation usa o serviço AWS Resource Access Manager (AWS RAM) para facilitar o compartilhamento de recursos. Quando você compartilha um recurso do catálogo com outra conta, AWS RAM envia um convite para a conta do beneficiário para aceitar ou rejeitar a concessão do recurso.

Serviços analíticos integrados, como o Amazon Athena e o Redshift Spectrum, exigem links de recursos para poder incluir recursos compartilhados nas consultas. Os diretores precisam criar um link de recurso AWS Glue Data Catalog para um recurso compartilhado de outra pessoa Conta da AWS. Para acessar mais informações sobre links de recursos, consulte Como os links de recursos funcionam em Lake Formation.

Um contêiner de link de catálogo é um objeto de catálogo de dados, que faz referência a um catálogo local ou de várias contas em nível de banco de dados federado de outras contas. AWS Você também pode criar links de banco de dados e links de tabela em um contêiner de links de catálogo. Ao criar um link de banco de dados ou um link de tabela, você deve especificar um recurso de destino que resida no mesmo catálogo de destino em nível de banco de dados do Amazon Redshift (banco de dados do Amazon Redshift).

Para criar um contêiner de links de catálogo, você precisa de `CREATE_CATALOG` do Lake Formation ou da permissão `glue:CreateCatalog`.

Você deve ter as configurações de versão da conta cruzada versão 4 ou superior para compartilhar bancos de dados ou tabelas no catálogo federado entre Conta da AWS s.

Criar um contêiner de links de catálogo para um catálogo federado entre contas

Você pode criar um contêiner de links de catálogo que aponta para um catálogo federado no nível do banco de dados do Redshift em qualquer AWS região usando o AWS Lake Formation console, AWS Glue `CreateCatalog` a API ou (). AWS Command Line Interface AWS CLI

Como criar um contêiner de links de catálogo para um catálogo compartilhado (console)

1. Abra o AWS Lake Formation console em https://console.aws.amazon.com/lakeformation/. Faça login como entidade principal com a permissão `CREATE_CATALOG` do Lake Formation.
2. No painel de navegação, escolha Catálogos e depois selecione Criar catálogo.
3. Na página Definir detalhes do catálogo, forneça as seguintes informações:

   Nome

   Digite um nome que siga as mesmas regras que um nome de catálogo. O nome pode ser o mesmo do catálogo compartilhado de destino.

   Tipo

   Escolha o contêiner de links de catálogo como o tipo de catálogo.

   Fonte

   Selecione `Redshift`.

   Catálogo de destino do Redshift

   Selecione um catálogo federado em nível de banco de dados do Redshift ou escolha um catálogo local (próprio) na lista.

   A lista contém todos os catálogos compartilhados com sua conta. Observe que o ID da conta do proprietário do catálogo está listado em cada catálogo. Caso você não veja um catálogo que saiba que foi compartilhado com sua conta, confira o seguinte:

   - Se você não for administrador do data lake, confira se o administrador do data lake concedeu a você as permissões do Lake Formation no catálogo.
   - Se você for administrador de um data lake e sua conta não estiver na mesma AWS Organization que a conta concedente, não se esqueça de aceitar o convite

de compartilhamento de recursos AWS Resource Access Manager (AWS RAM) para o catálogo. Para obter mais informações, consulte Aceitando um convite de compartilhamento de recursos do AWS RAM.

> **Note**
>
> Ao criar um contêiner de links de catálogo por meio do console, o menu suspenso Catálogo de destino do Redshift pode exibir `No matches` quando você tenta selecionar um catálogo do Redshift entre contas. Apesar dessa exibição, você pode inserir manualmente o ARN de destino do catálogo em nível de banco de dados federado do Amazon Redshift (banco de dados Amazon Redshift) no campo de entrada, e o formulário ainda funcionará corretamente. Por exemplo: `arn:aws:glue:us-east-1:123456789012:catalog/federated-catalog-redshift/dev`.
>
> Esse comportamento ocorre porque o console só pode pesquisar candidatos em potencial na conta atualmente conectada. O menu suspenso serve como um recurso de preenchimento automático, mas você ainda pode inserir ARNs manualmente para acesso entre contas.

4. Para permitir que os mecanismos de consulta Apache Iceberg leiam e gravem nos namespaces do Amazon Redshift, cria AWS Glue um cluster gerenciado do Amazon Redshift com os recursos computacionais e de armazenamento necessários para realizar operações de leitura e gravação sem afetar as cargas de trabalho do armazém de dados do Amazon Redshift. É preciso fornecer um perfil do IAM com as permissões necessárias para transferir dados de e para o bucket do Amazon S3.
5. Escolha Próximo.
6. (Opcional) Escolha Adicionar permissões para conceder permissões a outras entidades principais.

   No entanto, conceder permissões em um contêiner de link de catálogo não implica que as permissões serão concedidas no catálogo (vinculado) de destino. Você deve conceder permissões no catálogo de destino separadamente para que o link fique visível no Athena.
7. Depois, revise os detalhes do contêiner de links de catálogo e escolha Criar catálogo.

   Depois, você pode visualizar o nome do contêiner de links na página Catálogos.

Agora, você pode criar links de banco de dados e links de tabela no contêiner de links de catálogo para permitir o acesso por meio de mecanismos de consulta.

Criar um exemplo de CLI de contêiner de links de catálogo

- No exemplo a seguir, o objeto `TargetRedshiftCatalog` especifica o arn do catálogo em nível de banco de dados federado do Amazon Redshift (banco de dados do Amazon Redshift). O `DataLakeAccess` deve ser habilitado quando você cria o contêiner de links de catálogo.

```
aws glue create-catalog \
  --cli-input-json
    '{
        "Name": "linkcontainer",
        "CatalogInput": {
            "TargetRedshiftCatalog": {
               "CatalogArn": "arn:aws:us-east-1:123456789012:catalog/nscatalog/dev"
             },
            "CatalogProperties": {
              "DataLakeAccessProperties" : {
                "DataLakeAccess" : true,
                "DataTransferRole" : "arn:aws:iam::111122223333:role/
DataTransferRole"
             }
           }
        }
    }'
```

Criar links de recursos no contêiner de links de catálogo

É possível criar links de recursos para links de bancos de dados e tabelas em um contêiner de links de catálogo. Ao criar links de recursos de banco de dados ou de recursos de tabela, você deve especificar um recurso de destino que resida no mesmo catálogo de destino em nível de banco de dados do Amazon Redshift (banco de dados do Amazon Redshift) para o qual o contêiner de links aponta.

Você pode criar um link de recurso para um banco de dados compartilhado do Amazon Redshift ou uma tabela usando o AWS Lake Formation console, a API ou AWS Command Line Interface ()AWS CLI.

- Para obter instruções detalhadas, consulte Como criar um link de recurso para um banco de dados compartilhado do catálogo de dados.

  Veja a seguir um AWS CLI exemplo para criar um link de recurso de banco de dados em um contêiner de links de catálogo.

```
aws glue create-database \
  --cli-input-json \
    '{
        "CatalogId": "111122223333:linkcontainer",
        "DatabaseInput": {
            "Name": "dblink",
             "TargetDatabase": {
               "CatalogId": "123456789012:nscatalog/dev",
                "DatabaseName": "schema1"
             }
        }
    }'
```

- Para criar um link de recurso de tabela em um contêiner de link de catálogo, primeiro você precisa criar um AWS Glue banco de dados no local AWS Glue Data Catalog para conter o link de recurso de tabela.

  Para acessar mais informações sobre como criar links de recursos para tabelas compartilhadas, consulte Como criar um link de recurso para uma tabela compartilhada do catálogo de dados.

  - Criar um banco de dados para conter o exemplo de link de recurso de tabela

```
aws glue create-database \
  --cli-input-json \
      '{
          "CatalogId": "111122223333:linkcontainer",
          "DatabaseInput": {
              "Name": "db1",
              "Description": "creating parent database for table link"
          }
      }'
```

  - Exemplo de link de recurso de criação de tabela

```
aws glue create-table \
```

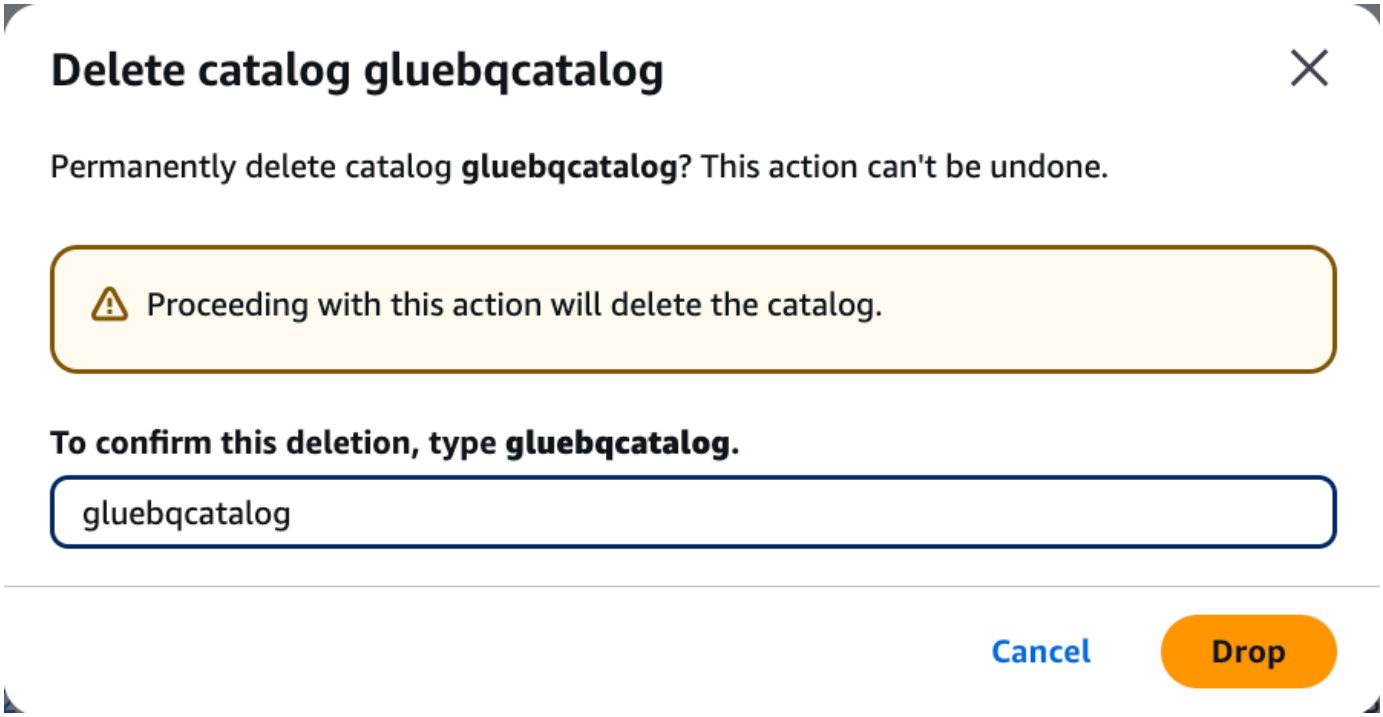
```
--cli-input-json \
  '{
      "CatalogId": "111122223333:linkcontainer",
       "DatabaseName": "db1",
      "TableInput": {
          "Name": "tablelink",
          "TargetTable": {
              "CatalogId": "123456789012:nscatalog/dev",
             "DatabaseName": "schema1",
              "Name": "table1"
           }
      }
  }'
```

Excluir um catálogo federado

Você pode excluir os catálogos federados que você criou AWS Glue Data Catalog usando a `glue:DeleteCatalog` operação ou o AWS Lake Formation console.

Como excluir um catálogo federado (console)

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.
2. No painel de navegação, escolha Catálogos, em Data Catalog.
3. Escolha o catálogo que deseja excluir da lista.
4. Escolha Excluir em Ações.
5. Escolha Eliminar para confirmar e o catálogo federado será excluído do Data Catalog.



Como excluir um catálogo federado (CLI)

- ```
  aws glue delete-catalog
   --catalog-id 123456789012:catalog name
  ```

Consultar catálogos federados

Depois de conceder permissões a outros diretores, eles podem fazer login e começar a consultar as tabelas nos catálogos federados fazendo login nas ferramentas SQL usando o Amazon Redshift, o Amazon EMR e o ETL. Amazon Athena AWS Glue

Para obter mais informações sobre como se conectar ao AWS Glue Data Catalog endpoint da extensão Apache Iceberg Rest ou ao aplicativo Spark independente, consulte Acessando a AWS Glue Data Catalog seção no Guia do desenvolvedor. AWS Glue

É possível usar as consultas da linguagem de definição de dados (DDL) para criar e gerenciar tabelas no banco de dados usando o Apache Spark no Amazon EMR. Para criar e excluir tabelas no banco de dados do Amazon Redshift, a entidade principal deve ter permissões `Create table` e `Drop` do Lake Formation.
  ```

Para acessar mais informações sobre como conceder permissões do Data Catalog, consulte Conceder permissões nos recursos do Catálogo de Dados.

Para obter mais informações sobre como consultar os recursos do catálogo Amazon Athena, consulte Consultando Amazon Athena no Guia AWS Glue Data Catalog do usuário do Amazon Athena.

Recursos adicionais do

Você pode usar o Amazon SageMaker Lakehouse para obter acesso unificado aos dados em data warehouses e data lakes. Por meio do SageMaker Lakehouse, você pode usar mecanismos preferenciais de análise, aprendizado de máquina e inteligência de negócios por meio de uma API REST aberta do Apache Iceberg para ajudar a garantir o acesso seguro aos dados com controles de acesso consistentes e refinados.

- SageMaker Workshop da Amazon
- Simplifique o acesso aos dados para sua empresa usando o Amazon SageMaker Lakehouse

Federando em fontes de dados externas no AWS Glue Data Catalog

Você pode conectar o AWS Glue Data Catalog (Catálogo de dados) a armazéns de dados, como Amazon Redshift, Snowflake, bancos de dados em nuvem, como Amazon RDS, Oracle, e serviços de streaming Amazon DynamoDB, como Amazon MSK, e sistemas locais, como Teradata, usando conexões. AWS Glue Essas conexões são armazenadas AWS Glue Data Catalog e registradas no AWS Lake Formation, permitindo que você crie um catálogo federado para cada fonte de dados disponível.

Catálogo federado é um contêiner de nível superior que aponta para um banco de dados em um sistema de dados externo. Ele permite que você consulte os dados diretamente do sistema de dados externo sem o processo de extração, transformação e carregamento (ETL).

Para obter mais informações sobre AWS Glue conexões, consulte Conexão com dados no Guia do AWS Glue desenvolvedor.

Os administradores do data lake podem criar catálogos federados usando o Amazon SageMaker Lakehouse ou. Amazon Athena

Os administradores do Data Lake podem então conceder permissões refinadas aos objetos dentro do catálogo usando o Lake Formation, controlando o acesso em vários níveis, como catálogo, banco de dados, tabela, coluna, linha ou célula. Os analistas de dados podem descobrir e consultar as fontes de dados catalogadas usando o Athena, com o Lake Formation aplicando as políticas de acesso definidas. Os analistas podem unir dados de várias fontes em uma única consulta sem precisar se conectar a cada fonte individualmente.

Tópicos

- Fluxo de trabalho
- Requisitos para conectar o Data Catalog a fontes de dados externas
- Criando um catálogo federado usando um AWS Glue gateway
- Visualizar objetos do catálogo
- Excluir um catálogo federado
- Consultar catálogos federados
- Recursos adicionais do

Fluxo de trabalho

Um administrador do data lake ou um usuário com as permissões necessárias conclui as etapas a seguir para conectá-lo AWS Glue Data Catalog a uma fonte de dados externa.

1. Cria uma AWS Glue conexão com a fonte de dados. Quando você registra a conexão, o perfil do IAM usado no registro da conexão deve ter acesso à função do Lambda e à localização do bucket de vazamento do Amazon S3.
2. Registra a conexão com o Lake Formation.
3. Cria um catálogo federado no Catálogo de Dados usando uma AWS Glue conexão para se conectar às fontes de dados disponíveis. Os bancos de dados, tabelas e visualizações são automaticamente catalogados no Data Catalog e registrados no Lake Formation.
4. Concede acesso a catálogos, bancos de dados e tabelas específicos para analistas de dados usando as permissões do Lake Formation. Fine-grained as políticas de controle de acesso podem ser definidas em lagos de dados, armazéns e fontes OLTP usando o Lake Formation, permitindo filtros de segurança em nível de linha e coluna.

   Os analistas de dados podem então acessar todos os dados por meio do Data Catalog usando consultas SQL no Athena, sem precisar de conexões ou credenciais de fonte de dados separadas.

Os analistas podem executar consultas SQL federadas que verificam dados de várias fontes, unindo dados no local sem pipelines de dados complexos.

Requisitos para conectar o Data Catalog a fontes de dados externas

Para AWS Glue Data Catalog conectar-se a fontes de dados externas, registrar a conexão com o Lake Formation e configurar catálogos federados, você precisa preencher os seguintes requisitos:

> **Note**
>
> Recomendamos que um administrador de data lake do Lake Formation crie as AWS Glue conexões para se conectar a fontes de dados externas e criar os catálogos federados.

1. Crie perfis do IAM.

   - Crie uma função que tenha as permissões necessárias para implantar os recursos (função Lambda, repositório de vazamento do Amazon S3, função do IAM e AWS Glue a conexão) necessários para criar uma conexão com a fonte de dados externa.
   - Crie uma função que tenha as permissões mínimas necessárias para acessar as propriedades da AWS Glue conexão (a função Lambda e o bucket de vazamento do Amazon S3). Esse é o perfil que você incluirá ao registrar a conexão no Lake Formation.

     Para usar o Lake Formation para gerenciar e proteger os dados em seu data lake, você deve registrar a AWS Glue conexão com o Lake Formation. Com isso, o Lake Formation pode fornecer credenciais ao Amazon Athena para consultar as fontes de dados federadas.

     O perfil deve ter as permissões `Select` ou `Describe` no bucket do Amazon S3 e na função do Lambda.

     - s3: ListBucket
     - s3: GetObject
     - lambda: InvokeFunction

     JSON

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

```
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3:*"
      ],
      "Resource": [
        "arn:aws:s3:::amzn-s3-demo-bucket1/object/*",
        "arn:aws:s3:::amzn-s3-demo-bucket1/object"
      ]
    },
    {
      "Sid": "lambdainvoke",
      "Effect": "Allow",
      "Action": "lambda:InvokeFunction",
      "Resource": "arn:aws:lambda:us-east-1:123456789012:function:example-
lambda-function"
    },
    {
      "Sid": "gluepolicy",
      "Effect": "Allow",
      "Action": "glue:*",
      "Resource": "*"
    }
  ]
}
```

- Adicione a política de confiança a seguir ao perfil do IAM que é usado para registrar a conexão:

  JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": [
                    "lakeformation.amazonaws.com"
              ]
            },
            "Action": "sts:AssumeRole"
```

```
        }
    ]
}
```

- O administrador do data lake que registra a conexão deve ter a permissão `iam:PassRole` no perfil.

  A seguir está uma política embutida que concede essa permissão. *<account-id>*Substitua por um número de AWS conta válido e *<role-name>* substitua pelo nome da função.

  JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "PassRolePermissions",
            "Effect": "Allow",
            "Action": [
                "iam:PassRole"
            ],
            "Resource": [
                "arn:aws:iam::111122223333:role/example-role-name>"
            ]
        }
    ]
}
```

- Para criar catálogos federados no Data Catalog, verifique se o perfil do IAM que você está usando é administrador de data lake do Lake Formation, conferindo as configurações do data lake (`aws lakeformation get-data-lake-settings`).

  Se você não for administrador de data lake, precisará da permissão `CREATE_CATALOG` do Lake Formation para criar um catálogo. O exemplo a seguir mostra como conceder as permissões necessárias para criar catálogos.

```
aws lakeformation grant-permissions \
--cli-input-json \
        '{
            "Principal": {
             "DataLakePrincipalIdentifier":"arn:aws:iam::123456789012:role/non-
admin"
```

```
            },
            "Resource": {
                "Catalog": {
                }
            },
            "Permissions": [
                "CREATE_CATALOG",
                "DESCRIBE"
            ]
        }'
```

Além da `CREATE_CATALOG` permissão, você precisa criar um catálogo federado `DATA_LOCATION_ACCESS` na conexão registrada. Um administrador de data lake tem permissões implícitas de localização de dados e não precisa dessa concessão. Para obter mais informações, consulte Permissões implícitas do Lake Formation.

O exemplo a seguir mostra como conceder `DATA_LOCATION_ACCESS` uma conexão registrada.

```
aws lakeformation grant-permissions \
  --cli-input-json \
    '{
      "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:role/non-admin-
role"
      },
      "Resource": {
        "DataLocation": {
          "CatalogId": "123456789012",
          "ResourceArn": "arn:aws:glue:us-east-1:123456789012:connection/
connection-name"
        }
      },
      "Permissions": ["DATA_LOCATION_ACCESS"]
    }'
```

2. Adicione a seguinte política de chaves à AWS KMS chave se você estiver usando uma chave gerenciada pelo cliente para criptografar os dados na fonte de dados. Substitua o número da conta por um número de AWS conta válido e especifique o nome da função. Por padrão, os dados são criptografados usando uma chave do KMS. O Lake Formation oferece uma opção

para criar sua chave do KMS personalizada para criptografia. Se você estiver usando uma chave gerenciada pelo cliente, deverá adicionar políticas de chave específicas à chave.

Para obter mais informações sobre como gerenciar as permissões de uma chave gerenciada pelo cliente, consulte Chaves gerenciadas pelo cliente.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "kms:Encrypt",
                "kms:Decrypt",
                "kms:ReEncrypt*",
                "kms:GenerateDataKey*",
                "kms:DescribeKey"
            ],
            "Resource": "arn:aws:kms:us-east-1:123456789012:key/key-1"
        }
    ]
}
```

Criando um catálogo federado usando um AWS Glue gateway

Para conectar-se AWS Glue Data Catalog às fontes de dados externas, você precisa usar AWS Glue conexões que permitam a comunicação com as fontes de dados externas. Você pode criar AWS Glue conexões usando o AWS Glue console, a API Create Connection e o console do Amazon SageMaker Lakehouse.

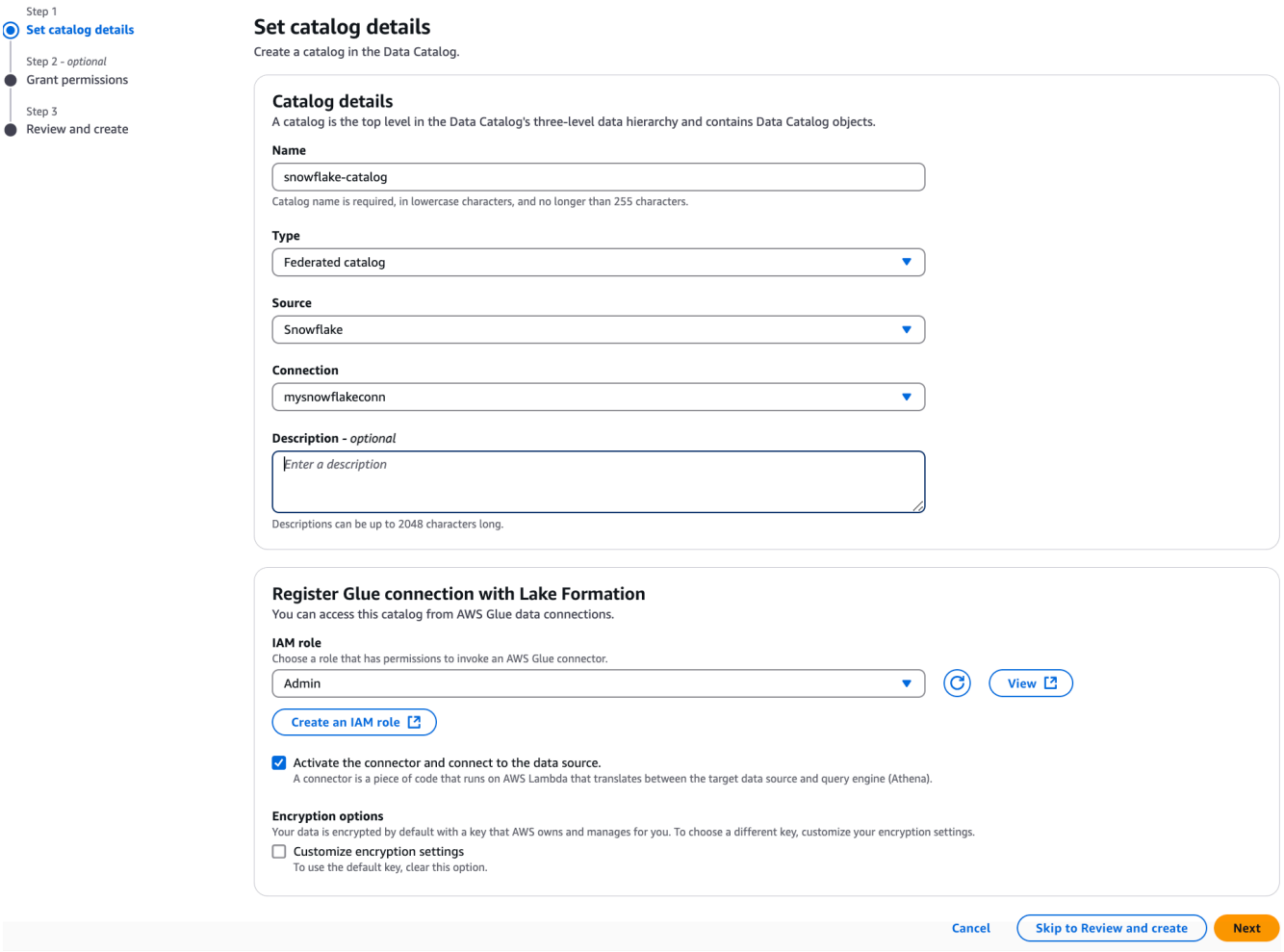
Para obter instruções passo a passo para criar uma AWS Glue conexão, consulte Conectando-se aos dados no Guia do AWS Glue desenvolvedor ou Criando conexões no Amazon SageMaker Lakehouse.

Quando um usuário executa uma consulta em tabelas federadas, o Lake Formation vende credenciais que invocam uma AWS Lambda função especificada na AWS Glue conexão para recuperar objetos de metadados da fonte de dados.

Console de gerenciamento da AWS

Como criar um catálogo federado por meio de uma fonte de dados externa e configurar permissões (console)

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.
2. No painel de navegação, escolha Catálogos, em Data Catalog.
3. Selecione a opção Criar catálogo.
4. Na página de detalhes Definir catálogo, insira as seguintes informações:



- Nome: um nome exclusivo para seu catálogo federado. O nome não pode ser alterado e deve estar em letras minúsculas. O nome pode ter no máximo 255 caracteres.
- Tipo: escolha o catálogo federado como o tipo de catálogo.
- Fonte: escolha uma fonte de dados no menu suspenso. As fontes de dados para as quais você criou conexões são exibidas. Para obter mais informações sobre como criar uma

AWS Glue conexão com uma fonte de dados externa, consulte Criação de conexões para conectores no Guia do AWS Glue desenvolvedor ou Criação de conexões no Amazon SageMaker Lakehouse.

- Conexão — Escolha uma AWS Glue conexão existente com a fonte de dados.
- Descrição: insira uma descrição para o catálogo criado por meio da fonte de dados.

5. Escolha um perfil do IAM a ser assumido pelo Lake Formation para fornecer credenciais para que o mecanismo de consulta acesse os dados da fonte de dados. Essa função deve ter as permissões necessárias para acessar a AWS Glue conexão e invocar a função Lambda para acessar dados da fonte de dados externa.

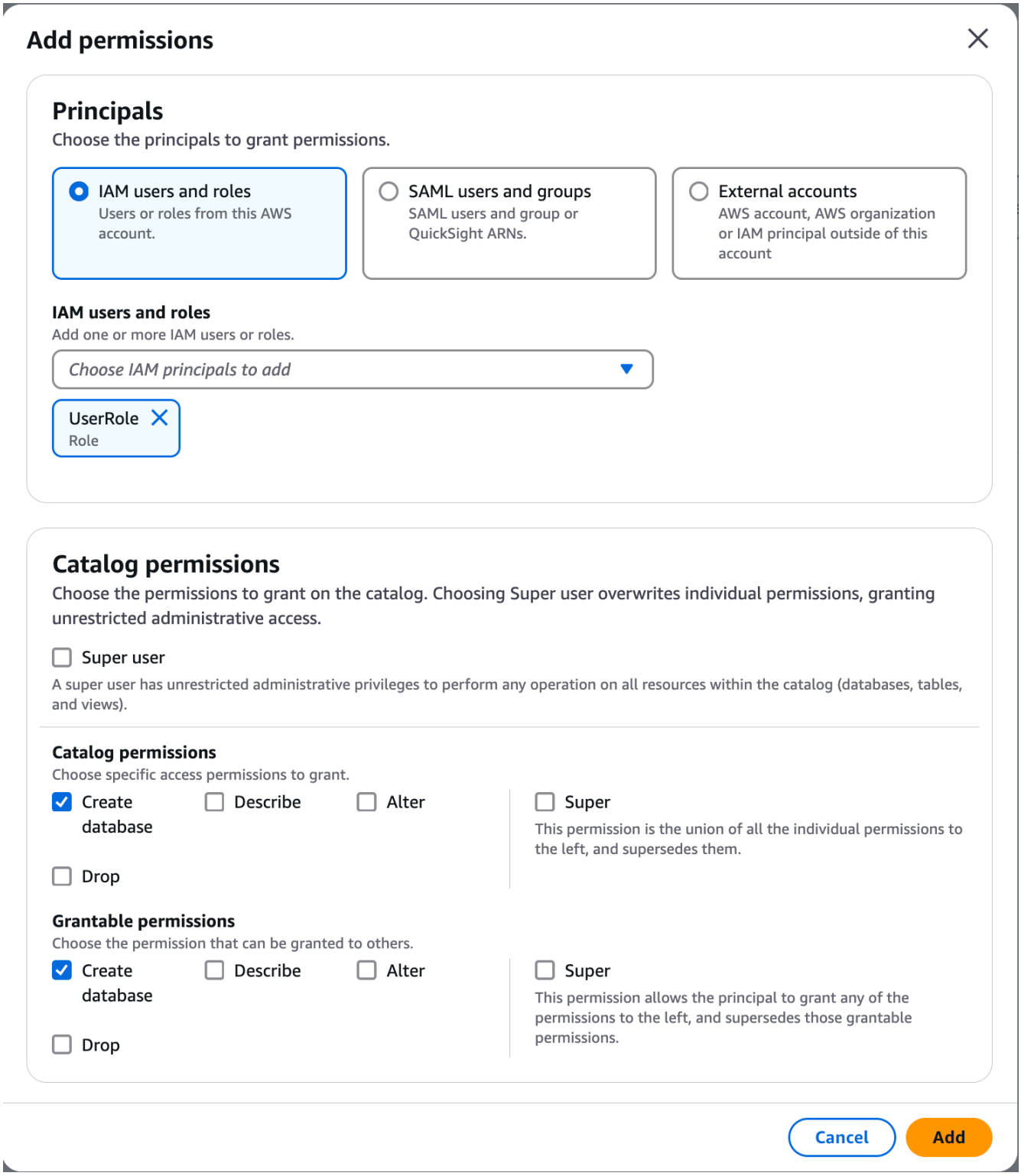
   Também é possível criar um perfil no console do IAM.

   Consulte a seção Requisitos para conectar o Data Catalog a fontes de dados externas para conhecer as permissões necessárias.

6. Selecione a opção Ativar o conector para se conectar à fonte de dados para permitir que o Athena execute consultas federadas.

   Para ver a lista de conectores aceitos, consulte Registrar sua conexão no Guia do usuário do Amazon Athena.

7. Opções de criptografia: escolha a opção Personalizar configurações de criptografia se quiser usar uma chave personalizada para criptografar o catálogo. Para usar uma chave personalizada, você deve adicionar outra política de chave gerenciada personalizada à sua chave do KMS.
8. Escolha Próximo para conceder permissões a outras entidades principais.
9. Na página Conceder permissões, escolha Adicionar permissões.
10. Na tela Adicionar permissões, escolha as entidades principais e os tipos de permissões a serem concedidas.



- Na seção Entidades principais, escolha um tipo de entidade principal e, em seguida, especifique as entidades principais para conceder permissões.

- Usuários e perfis do IAM: escolha um ou mais usuários ou perfis na lista de usuários e perfis do IAM.
- Usuários e grupos do SAML — Para SAML e Amazon Quick usuários e grupos, insira um ou mais Amazon Resource Names (ARNs) para usuários ou grupos federados por meio do SAML, ou ARNs para usuários ou grupos do Amazon Quick. Pressione Enter após cada ARN.

- Na seção Permissões, selecione permissões e permissões concedidas.

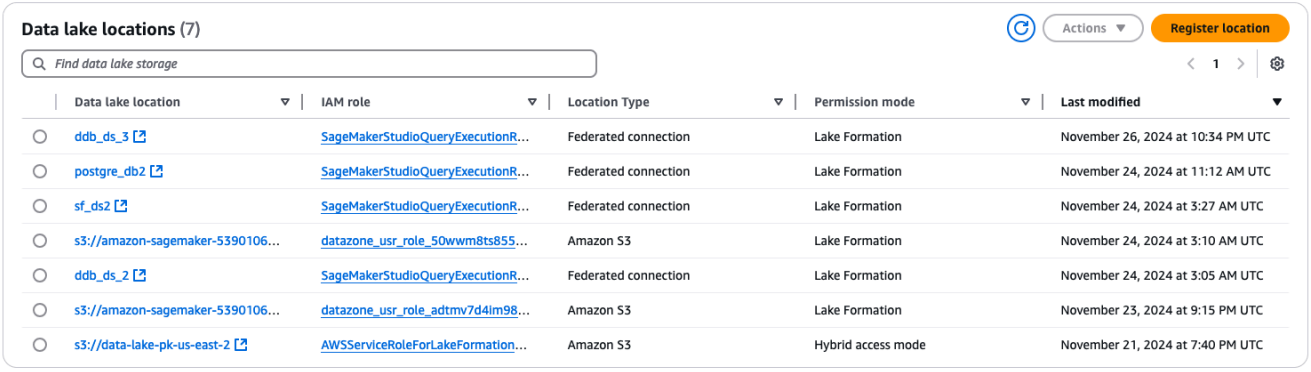
  Em Permissões do catálogo, selecione uma ou mais permissões a serem concedidas.

  Escolha Superusuário para conceder permissões administrativas irrestritas em todos os recursos do catálogo.

  Em Permissões concedidas, selecione as permissões que o beneficiário do subsídio pode conceder a outros diretores em sua conta. AWS Essa opção não é compatível quando você está concedendo permissões a uma entidade principal do IAM a partir de uma conta externa.

11. Escolha Próximo para revisar as informações e criar o catálogo. A lista Catálogos mostra o novo catálogo federado.

    A lista Locais de dados mostra a conexão federada recém-registrada.



AWS CLI

Como criar um catálogo federado por meio de uma fonte de dados externa e configurar permissões

1. O exemplo a seguir mostra como criar uma AWS Glue conexão.

```
aws glue create-connection
  --connection-input \
      '{
         "Name": "DynamoDB connection",
         "ConnectionType": "DYNAMODB",
         "Description": "A connection created for DynamoDB",
         "ConnectionProperties": {},
         "AthenaProperties": "spill_prefix": "your_spill_prefix",
         "lambda_function_arn": "Lambda_function_arn",
         "spill_bucket": "Your_Bucket_name",
         "AuthenticationConfiguration": {}
      }'
```

2. O exemplo a seguir mostra como registrar uma AWS Glue conexão com o Lake Formation.

```
aws lakeformation register-resource
  --cli-input-json \
    {"ResourceArn":"arn:aws:glue:us-east-1:123456789012:connection/
dynamo","RoleArn":"arn:aws:iam::123456789012:role/
AdminTelemetry","WithFederation":true}
```

3. Se você não for administrador do data lake, `DATA_LOCATION_ACCESS` conceda a conexão registrada ao principal que cria o catálogo federado. Um administrador de data lake tem permissões implícitas de localização de dados e não precisa dessa concessão. Para obter mais informações, consulte Permissões implícitas do Lake Formation.

   O exemplo a seguir mostra como conceder `DATA_LOCATION_ACCESS` uma conexão registrada.

```
aws lakeformation grant-permissions \
  --cli-input-json \
    '{
      "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:role/non-
admin-role"
      },
      "Resource": {
        "DataLocation": {
          "CatalogId": "123456789012",
```

```
          "ResourceArn": "arn:aws:glue:us-east-1:123456789012:connection/
connection-name"
        }
      },
      "Permissions": ["DATA_LOCATION_ACCESS"]
    }'
```

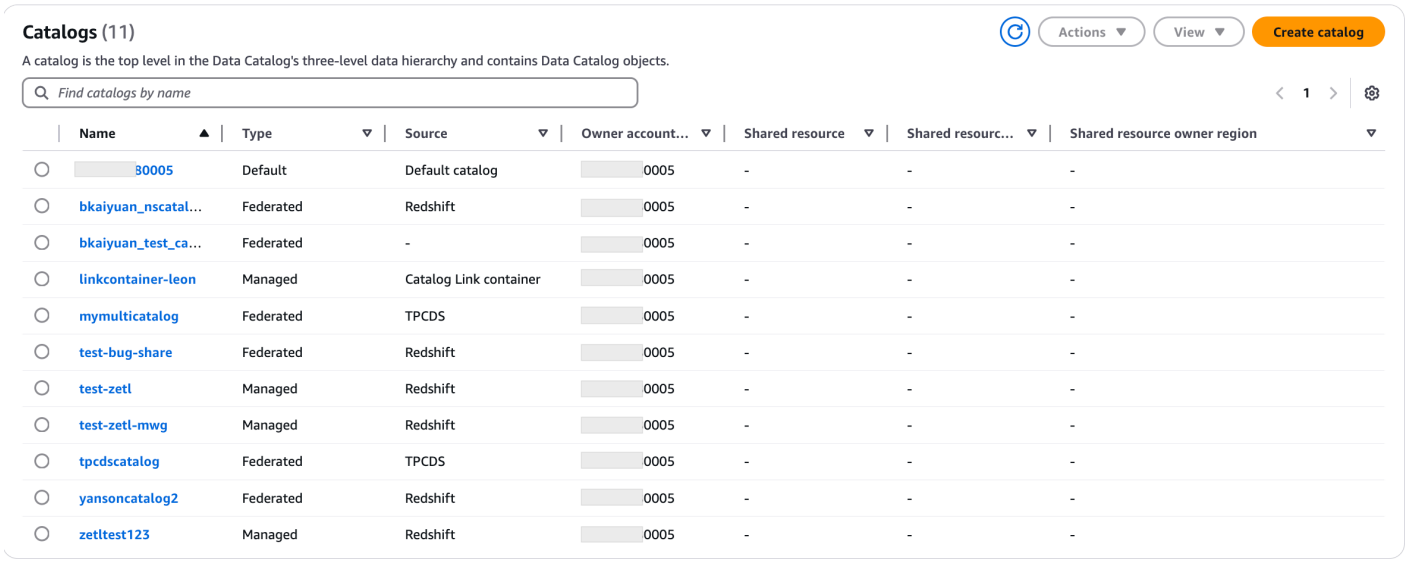
4. O exemplo a seguir mostra como criar um catálogo federado.

```
aws glue create-catalog
 --cli-input-json \
      '{
       "Name":"ddbcatalog",
       "CatalogInput":{"CatalogProperties":{"DataLakeAccessProperties":
{"DataTransferRole":"arn:aws:iam::123456789012:role/role name"}},
       "CreateDatabaseDefaultPermissions":[],
       "CreateTableDefaultPermissions":[],
       "FederatedCatalog":{"ConnectionName":"dynamo","Identifier":"dynamo"}
         }
       }'
```

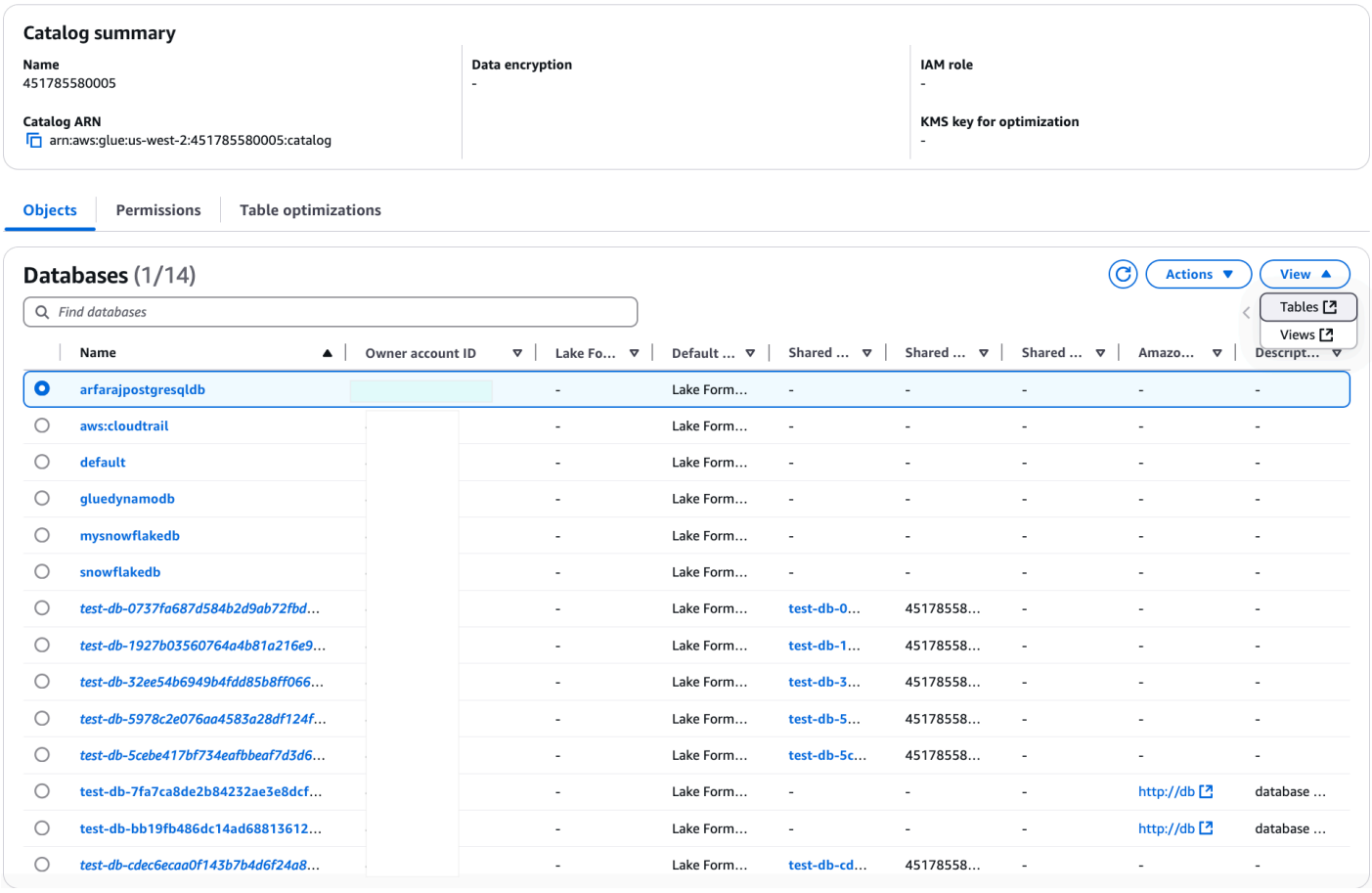
Visualizar objetos do catálogo

Para cada fonte de dados disponível, AWS Glue cria um catálogo correspondente no AWS Glue Data Catalog. Depois de criar um catálogo, você pode visualizar os objetos e tabelas nele usando o console do Lake Formation ou a AWS CLI. Para

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.
2. Escolha Catálogos no Data Catalog. A página de catálogos mostra os catálogos nos quais você tem permissões.



3. Escolha um catálogo na lista para visualizar os bancos de dados e tabelas contidos nele. A lista contém os bancos de dados em sua conta e os links de recursos, que são links para os bancos de dados e tabelas compartilhados em contas externas e são usados para acesso entre contas aos dados no data lake.



4. Escolha a opção Tabelas em Visualizar para visualizar e gerenciar as tabelas no banco de dados.

AWS CLI exemplos para visualizar catálogos e bancos de dados

O exemplo a seguir mostra como visualizar um catálogo usando AWS CLI

```
aws glue get-catalog \
--catalog-id 123456789012:dynamodbcatalog
```

O exemplo a seguir mostra como solicitar todos os catálogos na conta.

```
aws glue get-catalogs \
 --recursive
```

O exemplo de solicitação a seguir mostra como acessar os bancos de dados no catálogo.

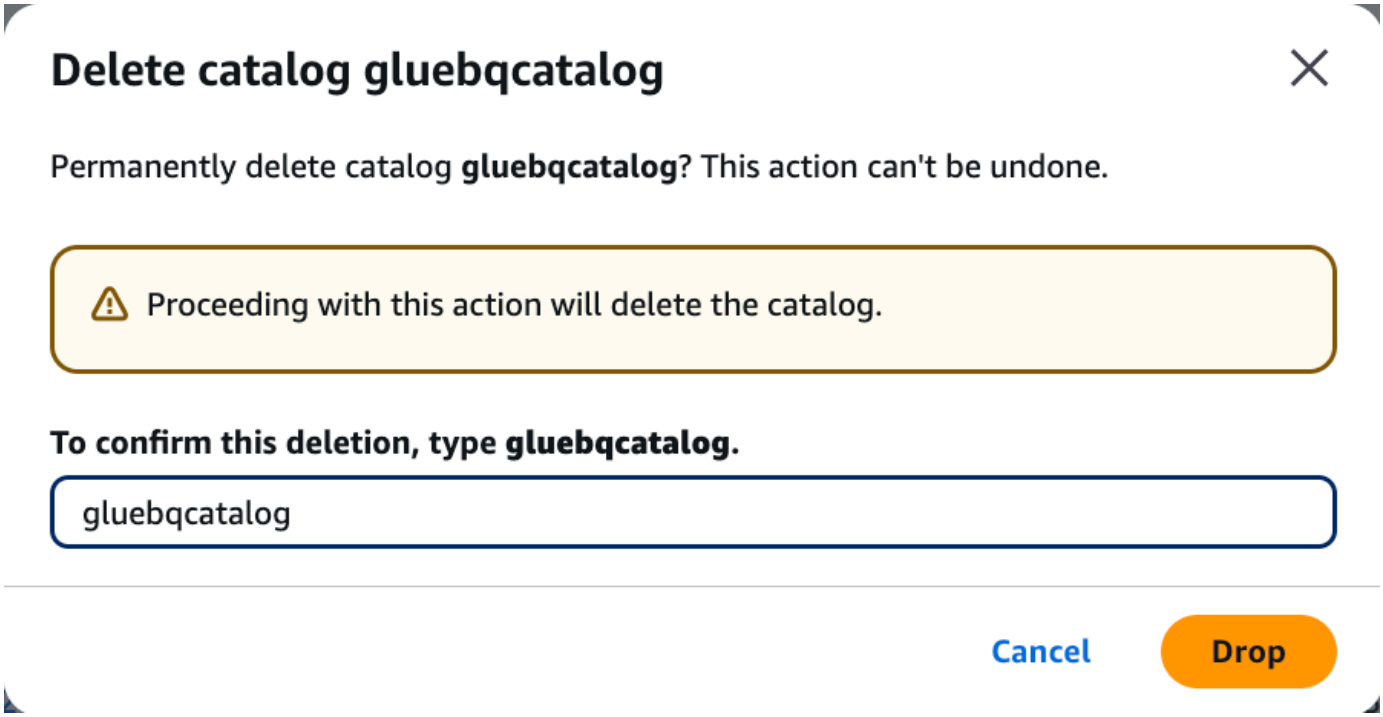
```
aws glue get-database \
--catalog-id 123456789012:dynamodbcatalog
--database-name database name
```

Excluir um catálogo federado

Você pode excluir os catálogos federados que você criou AWS Glue Data Catalog usando a `glue:DeleteCatalog` operação ou o AWS Lake Formation console.

Como excluir um catálogo federado (console)

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.
2. No painel de navegação, escolha Catálogos, em Data Catalog.
3. Escolha o catálogo que deseja excluir da lista.
4. Escolha Excluir em Ações.
5. Escolha Eliminar para confirmar e o catálogo federado será excluído do Data Catalog.



Como excluir um catálogo federado (CLI)

- ```
  aws glue delete-catalog
    --catalog-id 123456789012:catalog name
  ```

Consultar catálogos federados

Depois de conceder permissões a outras entidades principais, elas podem fazer login e começar a consultar as tabelas nos catálogos federados usando o Athena.

Para criar e excluir tabelas no banco de dados federado, a entidade principal deve ter as permissões `Create table` e `Drop` do Lake Formation.

Para acessar mais informações sobre como conceder permissões do Data Catalog, consulte Conceder permissões nos recursos do Catálogo de Dados.

Para obter mais informações sobre como consultar o catálogo de dados Amazon Athena, consulte Consultando Amazon Athena no Guia AWS Glue Data Catalog do usuário do Amazon Athena.
  ```

Recursos adicionais do

Nesta postagem do blog, mostramos como os analistas de dados agora podem acessar e consultar com segurança dados armazenados fora dos data lakes do S3, incluindo bancos de dados e Amazon DynamoDB bancos de dados do Amazon Redshift, por meio de uma experiência única e unificada. Agora, os administradores podem aplicar controles de acesso em diferentes níveis de granularidade para garantir que dados sensíveis permaneçam protegidos enquanto expandem o acesso aos dados. Isso permite que as organizações acelerem as iniciativas de dados enquanto mantêm a segurança e a conformidade, proporcionando uma tomada de decisão mais rápida e baseada em dados.

- Catalogue e controle as consultas federadas do Amazon Athena com o Amazon Lakehouse SageMaker

Integração de tabelas do Amazon S3 com AWS Glue Data Catalog and AWS Lake Formation

A funcionalidade Tabelas do Amazon S3 fornece armazenamento do S3 especificamente otimizado para workloads de analytics, melhorando a performance das consultas e reduzindo os custos. Os dados na funcionalidade Tabelas do S3 são armazenados em um novo tipo de bucket: um bucket de tabela, que armazena tabelas como sub-recursos. As tabelas do S3 têm suporte integrado para o padrão Apache Iceberg, que permite que você consulte facilmente dados em tabelas em buckets de tabelas do Amazon S3 usando mecanismos de consulta conhecidos, como o Apache Spark.

Você pode integrar as tabelas do Amazon S3 AWS Glue Data Catalog usando controles de acesso do IAM ou com subsídios do IAM e do Lake Formation:

- Controle de acesso do IAM: usa políticas do IAM para controlar o acesso às tabelas e ao catálogo de dados do S3. Nessa abordagem de controle de acesso, você precisa de permissões do IAM nos recursos das tabelas do S3 e nos objetos do catálogo de dados para acessar os recursos.
- Controle de acesso do Lake Formation: usa AWS Lake Formation concessões, além das permissões AWS Glue do IAM, para controlar o acesso às tabelas do S3 por meio do catálogo de dados. Nesse modo, os diretores exigem permissões do IAM para interagir com o catálogo de dados, e as concessões do Lake Formation determinam quais recursos do catálogo (bancos de dados, tabelas, colunas, linhas) o diretor pode acessar. Esse modo oferece suporte a controle de acesso de baixa granularidade (concessões em nível de banco de dados e em nível de tabela) e controle de acesso refinado (segurança em nível de coluna e nível de linha). Quando uma função registrada é configurada e a venda de credenciais está ativada, as permissões IAM do S3 Tables

não são necessárias para o diretor, pois o Lake Formation vende credenciais em nome do diretor usando a função registrada. O controle de acesso do Lake Formation também oferece suporte à venda de credenciais para mecanismos de análise de terceiros.

Esta seção fornece orientação para configurar a integração com AWS Lake Formation os seguintes cenários:

- Cenário A: você integrou tabelas e catálogo de dados do S3 usando controles de acesso do IAM e agora planeja usar AWS Lake Formation. Para saber mais, consulte Alterando os controles de acesso para integração com tabelas do S3.
- Cenário B: Você planeja integrar as tabelas e o catálogo de dados do S3 usando AWS Lake Formation e não os tem integrados em sua conta e região atualmente. Comece com a Pré-requisitos para integrar o catálogo de tabelas do Amazon S3 ao Data Catalog e o Lake Formation seção e sigaHabilitar a integração de Tabelas do Amazon S3.
- Cenário C: você integrou tabelas e catálogo de dados do S3 usando AWS Lake Formation e agora planeja usar o IAM. Para saber mais, consulte Alterando os controles de acesso para integração com tabelas do S3.

Certifique-se de seguir as etapas em Integração de tabelas do S3 com serviços de AWS análise para ter as permissões apropriadas para acessar os recursos AWS Glue Data Catalog e seus recursos de tabela e trabalhar com serviços de AWS análise.

Tópicos

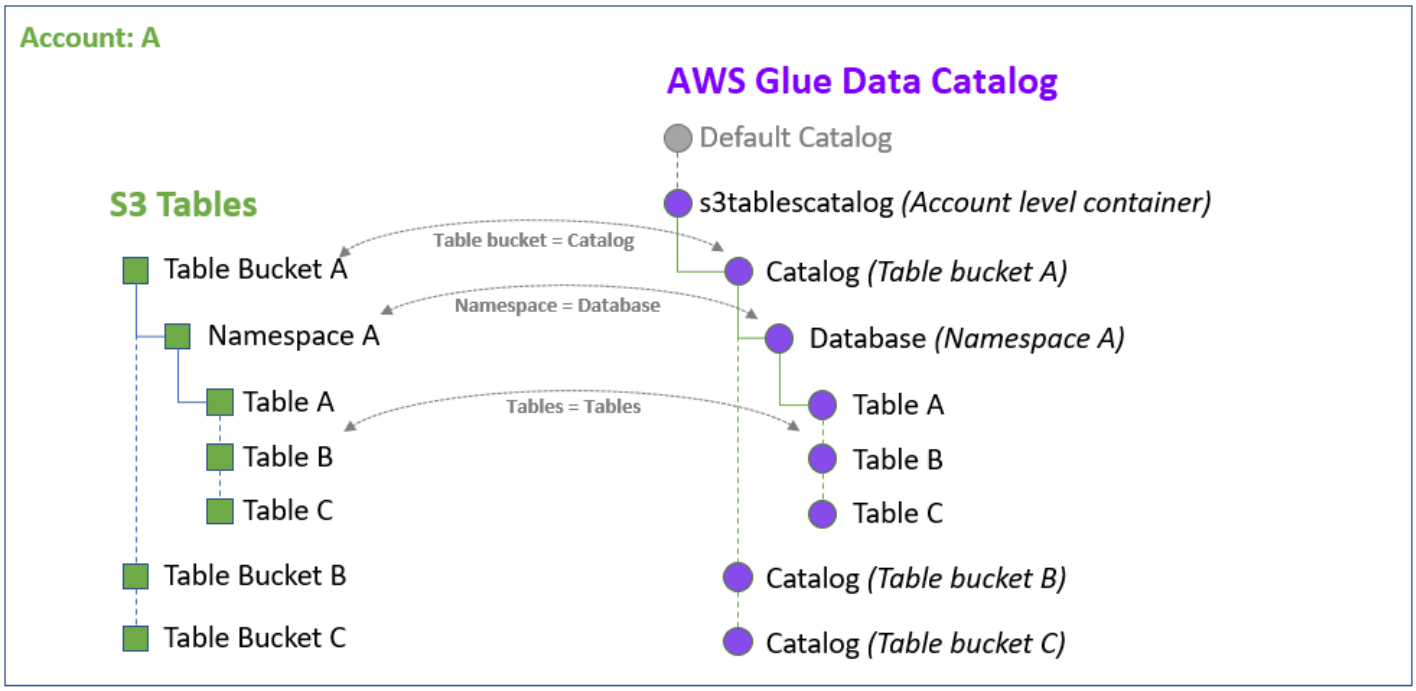
- Como funciona a integração do Data Catalog e do Lake Formation
- Pré-requisitos para integrar o catálogo de tabelas do Amazon S3 ao Data Catalog e o Lake Formation
- Habilitar a integração de Tabelas do Amazon S3
- Criar bancos de dados e tabelas no catálogo de Tabelas do S3
- Registrando um bucket de tabela do Amazon S3 em outro AWS account
- Conceder permissões

Como funciona a integração do Data Catalog e do Lake Formation

Quando você integra o catálogo de tabelas do S3 ao Data Catalog e ao Lake Formation, o serviço AWS Glue cria um único catálogo federado chamado `s3tablescatalog` no Data Catalog padrão

de sua conta específico da sua Região da AWS. A integração mapeia todos os recursos do bucket de tabelas do Amazon S3 em sua conta e Região da AWS no catálogo federado da seguinte maneira:

- Os buckets de tabela do Amazon S3 se tornam um catálogo de vários níveis no Data Catalog.
- O namespace do Amazon S3 associado é registrado como um banco de dados no Data Catalog.
- As tabelas do Amazon S3 no bucket de tabela se tornam tabelas no Data Catalog.



Após a integração com o Lake Formation, você pode criar tabelas Apache Iceberg no catálogo de tabelas e acessá-las por meio de mecanismos de AWS análise integrados, como o Amazon Athena Amazon EMR, bem como mecanismos de análise de terceiros.

Quando você também ativa o Lake Formation com integração, ele permite um controle de acesso refinado. AWS Lake Formation Essa abordagem de segurança significa que, além das permissões AWS Identity and Access Management (IAM), você deve conceder ao diretor do IAM as permissões do Lake Formation em suas tabelas antes de poder trabalhar com elas.

Há dois tipos principais de permissões no AWS Lake Formation:

- As permissões de acesso a metadados controlam a capacidade de criar, ler, atualizar e excluir tabelas e bancos de dados de metadados no Catálogo de Dados.

- As permissões de acesso aos dados subjacentes controlam a capacidade de ler e gravar dados nos locais do Amazon S3 subjacentes para os quais os recursos do Catálogo de Dados apontam.

O Lake Formation usa conjuntamente um modelo de permissões próprio e o modelo de permissões do IAM para controlar o acesso aos recursos do Catálogo de Dados e aos dados subjacentes:

- Para que uma solicitação de acesso aos recursos do Catálogo de Dados ou os dados subjacentes seja bem-sucedida, ela deve passar pelas verificações de permissão do IAM e do Lake Formation.
- As permissões do IAM controlam o acesso ao Lake Formation, às AWS Glue APIs e aos recursos, enquanto as permissões do Lake Formation controlam o acesso aos recursos do catálogo de dados, às localizações do Amazon S3 e aos dados subjacentes.

As permissões do Lake Formation se aplicam somente na região em que foram concedidas, e uma entidade principal deve ser autorizada por um administrador do data lake ou por outra entidade principal com as permissões necessárias para receber as permissões do Lake Formation.

Pré-requisitos para integrar o catálogo de tabelas do Amazon S3 ao Data Catalog e o Lake Formation

A seguir estão os pré-requisitos para habilitar a integração de tabelas do Amazon S3 com e. AWS Glue Data Catalog AWS Lake Formation

1. Quando você habilita a integração das tabelas do Amazon S3, o Lake Formation registra automaticamente o local das tabelas do S3. Para registrar a localização do bucket de tabelas no Lake Formation, você precisa de um IAM role/user com `lakeformation:RegisterResourcelakeformation:RegisterResourceWithPrivilegedAcc` e `lakeformation:CreateCatalog` permissões. Quando um usuário não administrador com essas permissões registra um local de catálogo, o Lake Formation concede automaticamente a ele a permissão `DATA_LOCATION_ACCESS` para esse local, possibilitando que a entidade principal da chamada tenha permissões para realizar todas as operações aceitas do Lake Formation no local de dados registrado.
2. Ao habilitar a integração de tabelas do S3, você precisa escolher um perfil do IAM para que o Lake Formation forneça credenciais a fim de permitir o acesso aos dados. Criar um perfil do IAM para acesso aos dados do Lake Formation aos buckets de tabelas do S3. O perfil do IAM usado para registrar o bucket de tabelas no Lake Formation requer as seguintes permissões:

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "LakeFormationPermissionsForS3ListTableBucket",
            "Effect": "Allow",
            "Action": [
                "s3tables:ListTableBuckets"
            ],
            "Resource": [
                "*"
            ]
        },
        {
            "Sid": "LakeFormationDataAccessPermissionsForS3TableBucket",
            "Effect": "Allow",
            "Action": [
                "s3tables:CreateTableBucket",
                "s3tables:GetTableBucket",
                "s3tables:CreateNamespace",
                "s3tables:GetNamespace",
                "s3tables:ListNamespaces",
                "s3tables:DeleteNamespace",
                "s3tables:DeleteTableBucket",
                "s3tables:CreateTable",
                "s3tables:DeleteTable",
                "s3tables:GetTable",
                "s3tables:ListTables",
                "s3tables:RenameTable",
                "s3tables:UpdateTableMetadataLocation",
                "s3tables:GetTableMetadataLocation",
                "s3tables:GetTableData",
                "s3tables:PutTableData"
            ],
            "Resource": [
                "arn:aws:s3tables:us-east-1:123456789012:bucket/*"
            ]
        }
    ]
```

```
}
```

Para obter mais informações, consulte Requisitos para funções usadas para registrar locais.

3. Adicione a política de confiança a seguir ao perfil do IAM para permitir que o serviço Lake Formation assuma o perfil e forneça credenciais temporárias aos mecanismos de analytics integrados.

```
{
  "Effect": "Allow",
  "Principal": {
    "Service": "lakeformation.amazonaws.com"
  },
  "Action": [
    "sts:AssumeRole",
    "sts:SetSourceIdentity",
    "sts:SetContext"  # add action to trust relationship when using IAM Identity
 center principals with Lake Formation
  ]
}
```

> **Note**
>
> O processo de integração dos serviços de AWS análise foi atualizado. Se você configurou a integração com a versão prévia, pode continuar usando sua integração atual. No entanto, o processo de integração atualizado oferece melhorias de desempenho. Para atualizar a integração:
>
> 1. Primeiro, exclua seu catálogo de tabelas do S3 existente no Lake Formation. Para excluir o catálogo, selecione o catálogo `s3tablescatalog` na lista de catálogos e escolha Excluir de Ações.
> 2. Depois, cancele o registro do local dos dados para o `s3tablescatalog`.
>   a. No console do Lake Formation, na seção Administrações, escolha Locais de dados.
>   b. Selecione um local e, no menu Ações, escolha Remover.
>   c. Quando solicitada a confirmação, escolha Remover.
>
>   Para obter instruções detalhadas sobre como cancelar o registro de um local de dados, consulte a seção. Cancelar o registro de uma localização do Amazon S3

3. Em seguida, siga as etapas de integração atualizadas na Habilitar a integração de Tabelas do Amazon S3 seção.

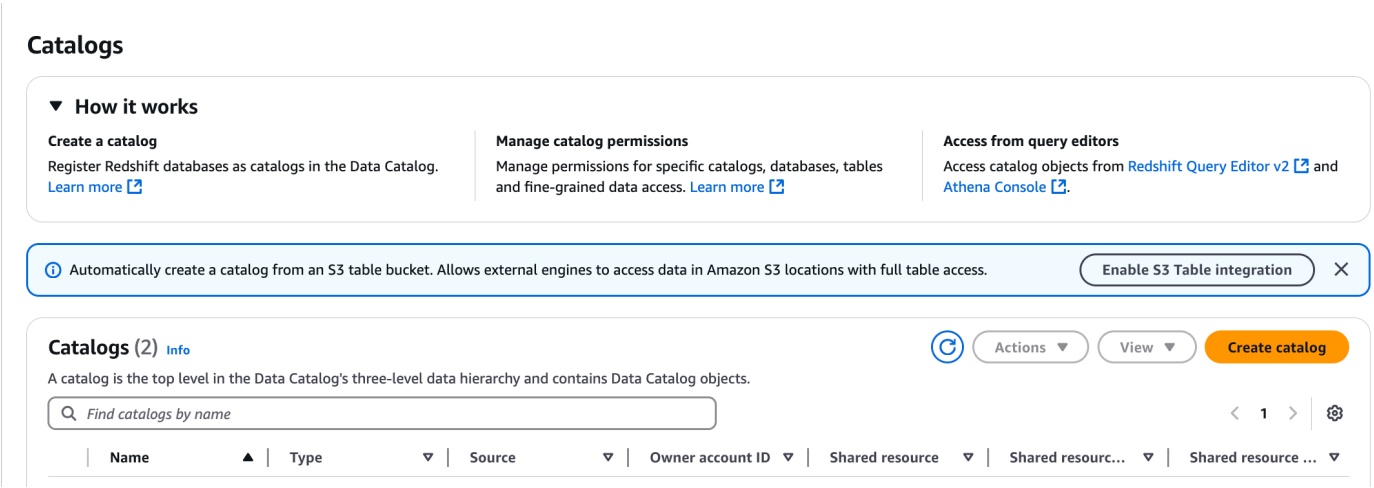
Habilitar a integração de Tabelas do Amazon S3

Você pode criar buckets de tabelas do Amazon S3 usando o console do Amazon S3 e integrá-los aos serviços de analytics da AWS . Para acessar mais informações, consulte Usar Tabelas do Amazon S3 com serviços de analytics da AWS.

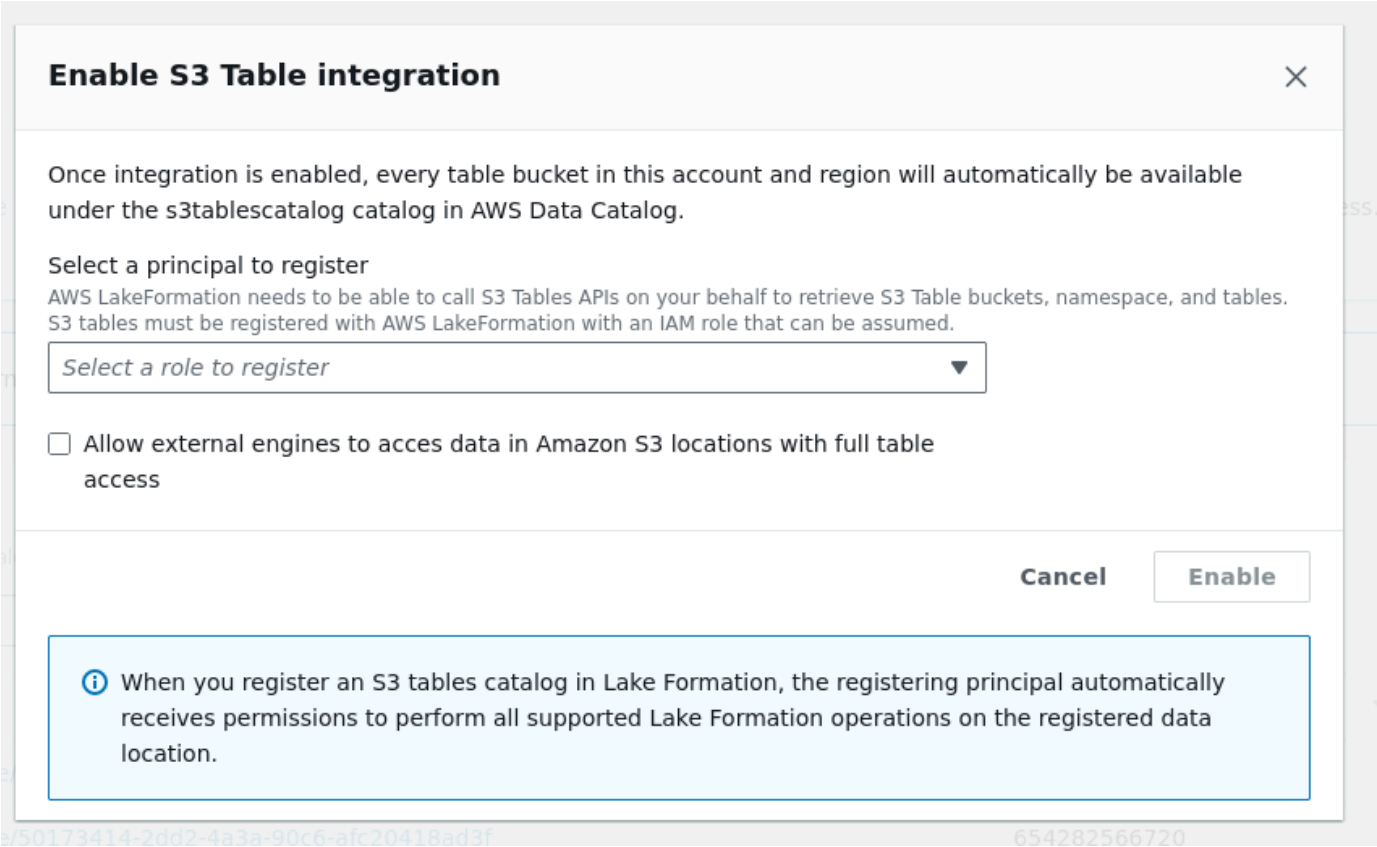
Em AWS Lake Formation, você pode habilitar a integração de tabelas do Amazon S3 com AWS Glue Data Catalog e AWS Lake Formation usando o console ou o uso do Lake Formation. AWS CLI

Como integrar as Tabelas do Amazon S3 ao Data Catalog e ao Lake Formation (console)

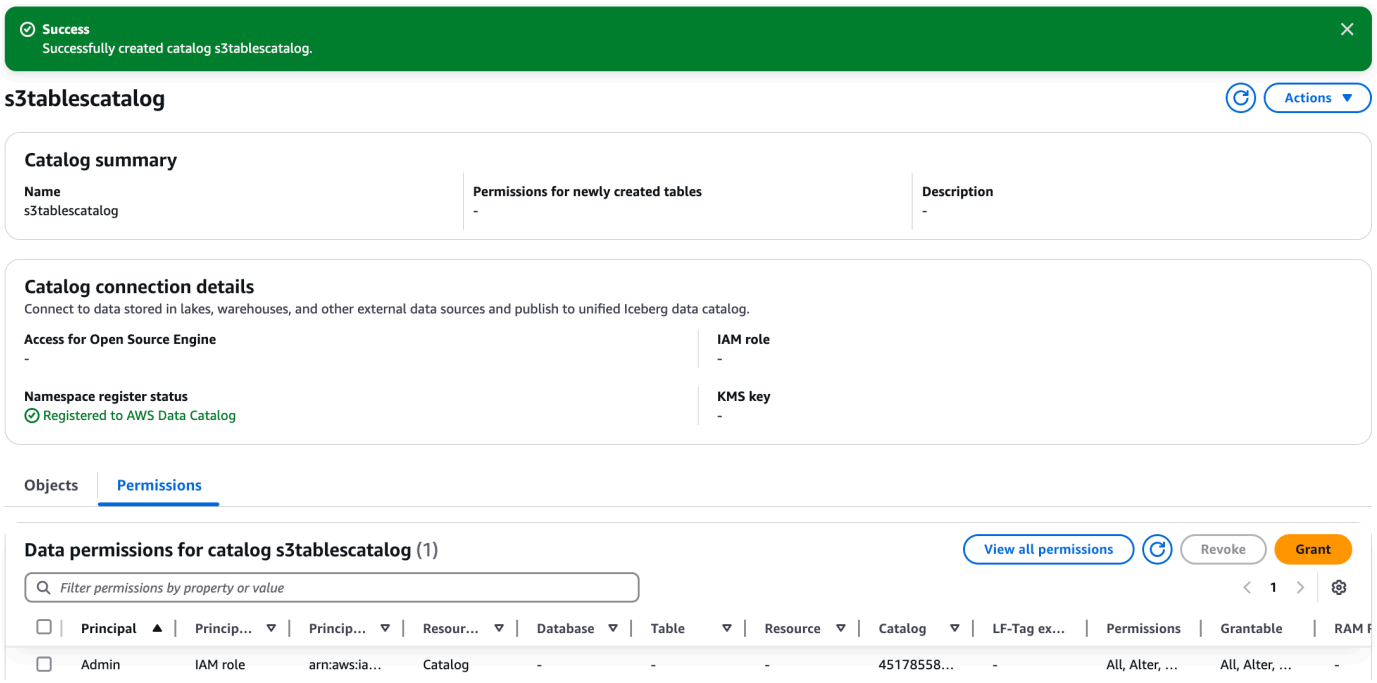
1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.
2. No painel de navegação, escolha Catálogos, em Data Catalog.
3. Escolha Habilitar integração de Tabelas do S3 na página Catálogos.



4. Escolha um perfil do IAM com as permissões necessárias para que o Lake Formation assuma o fornecimento de credenciais aos mecanismos de consulta analítica. Para ter as permissões necessárias para que o perfil acesse dados, consulte step3-permissions na seção de pré-requisitos.



5. Selecione a opção Permitir que mecanismos externos acessem dados em locais do Amazon S3 com acesso total à tabela. Quando você habilita o acesso total à tabela para mecanismos de terceiros, o Lake Formation exibe as credenciais diretamente para o mecanismo de terceiros, sem realizar a validação da tag de sessão do IAM. Isso significa que você não pode aplicar controles de acesso refinados do Lake Formation às tabelas que estão sendo acessadas.
6. Escolha Habilitar. O novo catálogo para Tabelas do S3 é adicionado à lista de catálogos. Quando você habilita a integração do catálogo de Tabelas do S3, o serviço registra o local dos dados do bucket de Tabelas do S3 no Lake Formation.
7. Escolha o catálogo para visualizar os objetos e conceder permissões a outras entidades principais.



Para criar catálogos de vários níveis, consulte a seção Criar um bucket de tabelas no Guia do usuário do Amazon Simple Storage Service.

Como integrar as Tabelas do Amazon S3 ao Data Catalog e ao Lake Formation (CLI)

Seguindo a seção de pré-requisitos, crie uma função de serviço do IAM que permita que o Lake Formation acesse seus recursos de tabela.

1. Crie um arquivo chamado `Role-Trust-Policy.json` que contenha a seguinte política de confiança:

```
{
    "Version": "2012-10-17",
    "Statement": [
      {
        "Sid": "LakeFormationDataAccessPolicy",
        "Effect": "Allow",
        "Principal": {
          "Service": "lakeformation.amazonaws.com"
        },
        "Action": [
            "sts:AssumeRole",
            "sts:SetContext",
            "sts:SetSourceIdentity"
```

```
        ],
        "Condition": {
          "StringEquals": {
            "aws:SourceAccount": "111122223333"
          }
        }
      }
    ]
}
```

2. Crie um perfil de serviço do IAM usando o seguinte comando:

```
aws iam create-role \
  --role-name S3TablesRoleForLakeFormation \
  --assume-role-policy-document file://Role-Trust-Policy.json
```

3. Crie um arquivo chamado `LF-GluePolicy.json` que contenha a seguinte política:

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "LakeFormationPermissionsForS3ListTableBucket",
            "Effect": "Allow",
            "Action": [
                "s3tables:ListTableBuckets"
            ],
            "Resource": [
                "*"
            ]
        },
        {
            "Sid": "LakeFormationDataAccessPermissionsForS3TableBucket",
            "Effect": "Allow",
            "Action": [
                "s3tables:CreateTableBucket",
                "s3tables:GetTableBucket",
                "s3tables:CreateNamespace",
                "s3tables:GetNamespace",
                "s3tables:ListNamespaces",
                "s3tables:DeleteNamespace",
                "s3tables:DeleteTableBucket",
                "s3tables:CreateTable",
```

```
                "s3tables:DeleteTable",
                "s3tables:GetTable",
                "s3tables:ListTables",
                "s3tables:RenameTable",
                "s3tables:UpdateTableMetadataLocation",
                "s3tables:GetTableMetadataLocation",
                "s3tables:GetTableData",
                "s3tables:PutTableData"
            ],
            "Resource": [
                "arn:aws:s3tables:us-east-1:111122223333:bucket/*"
            ]
        }
    ]
}
```

4. Anexe a política ao perfil usando o seguinte comando:

```
aws iam put-role-policy \
  --role-name S3TablesRoleForLakeFormation \
  --policy-name LakeFormationDataAccessPermissionsForS3TableBucket \
  --policy-document file://LF-GluePolicy.json
```

5. Crie um arquivo chamado `input.json` que contenha o seguinte:

```
{
    "ResourceArn": "arn:aws:s3tables:us-east-1:111122223333:bucket/*",
    "WithFederation": true,
    "RoleArn": "arn:aws:iam::111122223333:role/S3TablesRoleForLakeFormation"
}
```

6. Registre os buckets de tabela no Lake Formation usando o seguinte comando:

```
aws lakeformation register-resource \
  --region us-east-1 \
  --with-privileged-access \
  --cli-input-json file://input.json
```

7. Crie um arquivo chamado `catalog.json` que contenha o seguinte catálogo:

```
{
   "Name": "s3tablescatalog",
   "CatalogInput": {
```

```
      "FederatedCatalog": {
          "Identifier": "arn:aws:s3tables:us-east-1:111122223333:bucket/*",
          "ConnectionName": "aws:s3tables"
       },
       "CreateDatabaseDefaultPermissions": [],
       "CreateTableDefaultPermissions": [],
       "AllowFullTableExternalDataAccess": "True"
   }
}
```

8. Crie o catálogo `s3tablescatalog` usando o comando a seguir. A criação desse catálogo preenche o AWS Glue Data Catalog com objetos correspondentes a compartimentos de tabela, namespaces e tabelas.

```
aws glue create-catalog \
  --region us-east-1 \
  --cli-input-json file://catalog.json
```

9. Verifique se o `s3tablescatalog` catálogo foi adicionado AWS Glue usando o seguinte comando:

```
aws glue get-catalog --catalog-id s3tablescatalog
```

Criar bancos de dados e tabelas no catálogo de Tabelas do S3

Você pode criar bancos de dados para organizar suas tabelas do Apache Iceberg e tabelas para definir o esquema e o local dos seus dados no catálogo de Tabelas do S3.

Criar um banco de dados (Console)

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/e faça login como administrador do data lake ou criador de banco de dados.
2. No painel de navegação, selecione Bancos de dados, em Catálogo de dados.
3. Selecione Criar banco de dados.
4. Na página Criar banco de dados, escolha a opção Banco de dados e insira os seguintes detalhes:

   - Nome: um nome exclusivo para o banco de dados.

- Catálogo de dados: escolha o catálogo de Tabelas do S3. O banco de dados residirá nesse catálogo.
- Descrição: (opcional) adicione uma descrição e um local.
- Controle de acesso do IAM para novas tabelas: opcionalmente, selecione Usar somente controle de acesso do IAM para novas tabelas nesse banco de dados. Para acessar informações sobre essa opção, consulte a seção Alterar as configurações padrão do seu data lake.
- Escolha Criar banco de dados. Você pode ver o banco de dados criado no catálogo de Tabelas do S3.

Crie um banco de dados usando AWS CLI

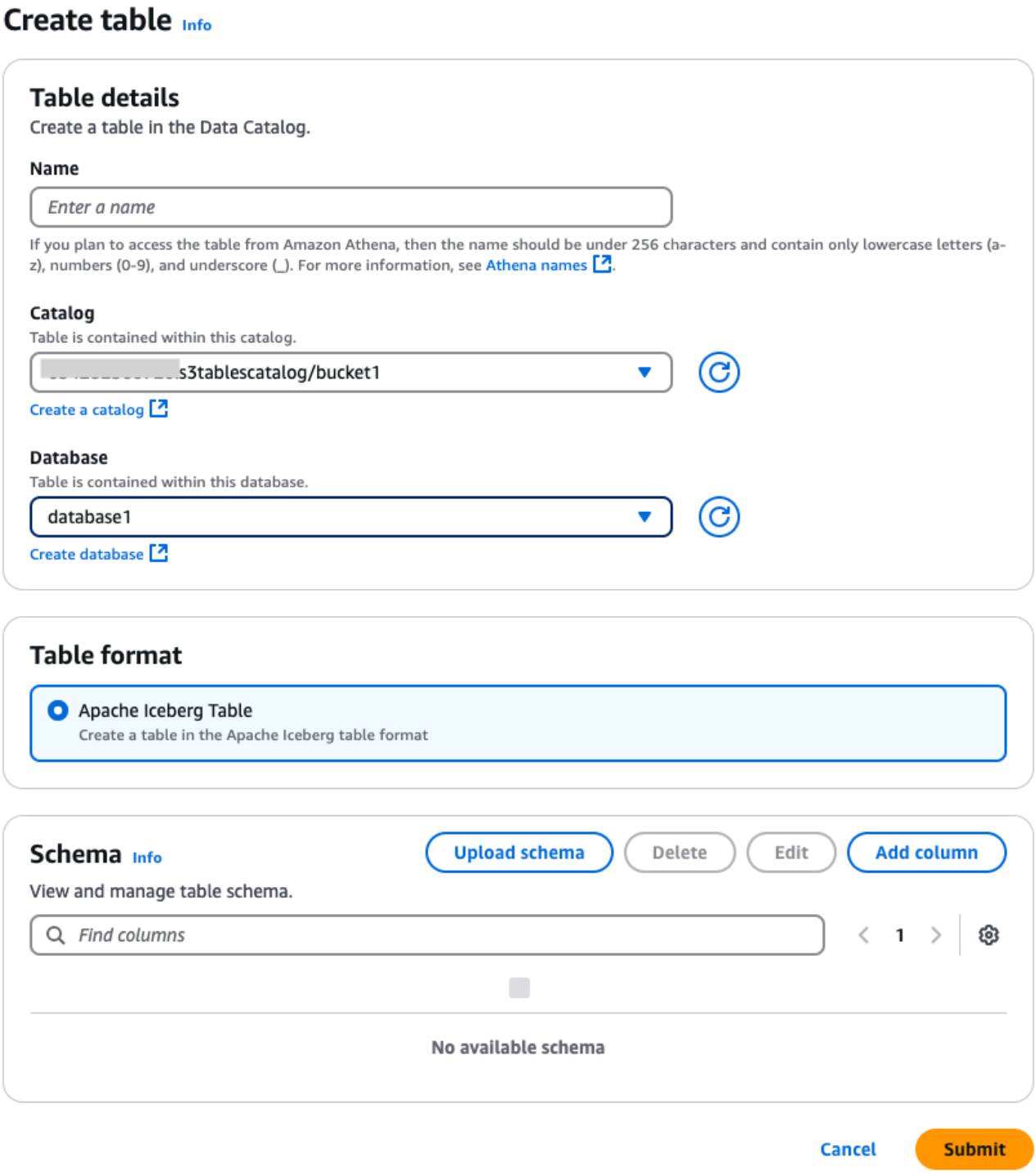
O comando da CLI a seguir mostra como criar um banco de dados no catálogo de Tabelas do S3.

```
aws glue create-database
--region us-east-1 \
--catalog-id "123456789012:s3tablescatalog/test" \
--database-input \
 '{ "Name": "testglueclidbcreation" }'
```

Crie uma tabela (Console de gerenciamento da AWS)

Você pode criar tabelas de metadados do Apache Iceberg no catálogo de tabelas do S3 usando o console do Lake Formation ou a API. AWS Glue `CreateTable`

1. Abra o console do Lake Formation em e faça login como administrador do data lake ou usuário com `CreateTable` permissão. https://console.aws.amazon.com/lakeformation/
2. No painel de navegação, escolha Tabelas no Data Catalog.
3. Escolha Criar tabela.
4. Na página Criar tabela, insiraos detalhes da tabela:



- Nome: insira um nome exclusivo para a tabela.
- Catálogo: escolha o catálogo de Tabelas do S3 como catálogo.
- Banco de dados: escolha o banco de dados no catálogo de Tabelas do S3.

- Descrição: insira uma descrição para a tabela.
- Esquema: escolha Adicionar colunas para adicionar colunas e tipos de dados das colunas. Você tem a opção de criar uma tabela vazia e atualizar o esquema posteriormente. O Iceberg permite que você evolua o esquema e a partição depois de criar a tabela. Você pode usar as consultas do Athena para atualizar o esquema da tabela e as consultas do Spark para atualizar as partições.

5. Selecione Enviar.

Crie uma tabela (AWS CLI)

```
aws glue create-table \
--database-name "testglueclidbcreation" \
--catalog-id "123456789012:s3tablescatalog/test" \
--region us-east-1 \
--table-input \
'{ "Name": "testtablegluecli", "Parameters": { "format": "ICEBERG" },
 "StorageDescriptor": { "Columns": [ {"Name": "x", "Type": "int", "Parameters":
 {"required": "true"}} ] } }'
```

Registrando um bucket de tabela do Amazon S3 em outro AWS account

Você pode registrar buckets de mesa individuais do Amazon S3 (formato ARN:`arn:aws:s3tables:us-east-1:`*`account-id`*`:bucket/`*`bucket-name`*) de uma conta com AWS Lake Formation em outra conta. Por exemplo, você pode registrar um bucket de tabelas da conta A no Lake Formation da conta B.

Pré-requisitos

Antes de iniciar o registro do bucket entre contas:

- Crie um bucket de tabela na conta A.
- Crie um perfil do IAM na conta B com as permissões apropriadas para o registro do bucket.

  Para acessar mais informações sobre as permissões necessárias para registrar um bucket de tabelas no Lake Formation, consultePré-requisitos para integrar o catálogo de tabelas do Amazon S3 ao Data Catalog e o Lake Formation.
- Registre o bucket de tabelas na conta em que você criará o catálogo de Tabelas do S3.

- Para acesso entre contas, o perfil especificado ao registrar o bucket de tabelas deve ser um perfil na conta com as permissões apropriadas para acessar o bucket entre contas. O perfil precisa das ações do IAM das Tabelas do S3 necessárias para acessar os recursos do bucket.

Cross-account registro de balde de mesa

Nos procedimentos a seguir, a conta A é a conta proprietária do recurso e a conta B é onde o bucket de tabelas será registrado para gerenciar as permissões de acesso.

1. Faça login Console de gerenciamento da AWS na conta A.

   Abra o console do Amazon S3 em https://console.aws.amazon.com/s3/.

2. Crie um bucket de tabela. Para acessar mais informações, consulte Criar um bucket de tabela no Guia do usuário do Amazon S3.

3. Registre o bucket de tabelas na conta B.

   Use o AWS CLI para registrar o balde de mesa da conta A com Lake Formation na conta B.

   ```
   aws lakeformation register-resource \
   --resource-arn 'arn:aws:s3tables:us-east-1:account-A-id:bucket/single-bucket-name'
    \
   --role-arn arn:aws:iam::account-B-id:role/role-name \
   --region us-east-1
   ```

   Substitua account-A-id, single-bucket-name, account- B-id e role-name pelos seus valores específicos.

4. Depois, crie um catálogo para o bucket de tabelas na conta B.

   Crie um catálogo usando a AWS CLI.

   ```
   aws glue create-catalog --region us-east-1 \
   --cli-input-json \
   '{
      "Name": "catalog-name",
      "CatalogInput" : {
         "FederatedCatalog": {
            "Identifier": "arn:aws:s3tables:us-east-1:account A:bucket/single-bucket-
   name",
            "ConnectionName": "aws:s3tables"
         },
   ```

```
      "CreateDatabaseDefaultPermissions": [],
      "CreateTableDefaultPermissions": []
   }
}'
```

Substitua catalog-name, account- A-id e bucket-name por seus valores específicos.

O exemplo de CLI a seguir mostra como visualizar os detalhes do catálogo.

```
aws glue get-catalog \
  --catalog-id account-id:catalog-name \
  --region us-east-1
```

5. Depois, crie bancos de dados e tabelas no catálogo recém-criado na conta B.

   Crie um banco de dados.

```
aws glue create-database \
  --region us-east-1 \
  --catalog-id "account-B-id:catalog-name" \
  --database-input \
'{
  "Name": "database-name"
}'
```

   Crie uma tabela.

```
aws glue create-table \
  --database-name database-name \
  --catalog-id account-B-id:catalog-name\
  --region us-east-1 \
  --table-input \
  '{
        "Name": "table-name",
        "Parameters": {
            "format": "ICEBERG"
        },
        "StorageDescriptor": {
           "Columns": [
        {"Name": "x", "Type": "int", "Parameters": {"required": "true"}}
          ]
```

```
        }
}'
```

Substitua nome do banco de dados, nome da contaB-id, nome do catálogo e nome da tabela por seus valores específicos.

6. Os exemplos a seguir mostram como visualizar os objetos no catálogo.

   Visualize detalhes do banco de dados.

   ```
   aws glue get-database \
     --name database-name \
     --catalog-id account-B-id:catalog-name \
     --region us-east-1
   ```

   Visualize os detalhes da tabela.

   ```
   aws glue get-table \
     --name table-name \
     --database-name database-name \
     --catalog-id account-B-id:catalog-name \
     --region us-east-1
   ```

   Substitua nome do banco de dados, nome da contaB-id, nome do catálogo e nome da tabela por seus valores específicos.

Conceder permissões

Depois de integrar suas tabelas do S3 AWS Lake Formation, você pode conceder permissões no catálogo de tabelas do S3 e nos objetos do catálogo (compartimentos de tabelas, bancos de dados, tabelas) para outras funções e usuários do IAM em sua conta. As permissões do Lake Formation permitem definir controles de acesso em granularidade em nível de tabela, coluna e linha para usuários de mecanismos analíticos integrados, como o Amazon Redshift Spectrum e o Athena.

Você pode conceder permissões usando o método de recurso nomeado ou o método de controle de acesso (LF-TBAC) baseado na tag Lake Formation. Antes de conceder permissões de uso LF-Tags de LF-Tag expressões, você deve defini-las e atribuí-las aos objetos do Catálogo de Dados.

Para obter mais informações, consulte Gerenciamento do controle LF-Tags de acesso aos metadados.

Você pode compartilhar bancos de dados e tabelas com AWS contas externas concedendo permissões do Lake Formation às contas externas. Os usuários podem então executar consultas e trabalhos que unem e consultam tabelas em várias contas. Quando você compartilha um recurso do catálogo com outra conta, as entidades principais dessa conta podem operar nesse recurso como se ele estivesse no respectivo Data Catalog.

Quando você compartilha bancos de dados e tabelas com contas externas, a permissão de superusuário não está disponível.

Para receber instruções detalhadas sobre como conceder permissões, consulte a seção Gerenciando permissões do Lake Formation.

AWS CLI exemplo de concessão de permissões em uma tabela do Amazon S3

```
aws lakeformation grant-permissions \
--cli-input-json \
'{
    "Principal": {
        "DataLakePrincipalIdentifier":"arn:aws:iam::111122223333:role/DataAnalystRole"
    },
    "Resource": {
        "Table": {
            "CatalogId":"111122223333:s3tablescatalog/amzn-s3-demo-bucket1",
            "DatabaseName":"S3 table bucket namespace <example_namespace>",
            "Name":"S3 table bucket table name <example_table>"
        }
    },
    "Permissions": [
        "SELECT"
    ]
}'
```

Os seguintes parâmetros devem ser inseridos no comando:

- DataLakePrincipalIdentifier — ARN de usuário, função ou grupo do IAM para conceder permissões
- CatalogId — ID da AWS conta de 12 dígitos que possui o Catálogo de Dados
- DatabaseName — Nome do namespace do bucket de tabelas do Amazon S3
- Nome: nome da tabela do bucket de Tabelas do Amazon S3.

- Permissões: permissões a serem concedidas. As opções são: SELECT, INSERT, DELETE, DESCRIBE, ALTER, DROP, ALLL e SUPER.

Acessar Tabelas do Amazon S3 compartilhadas

Depois de conceder permissões entre contas em um banco de dados ou tabela no catálogo de Tabelas do S3, para acessar os recursos, você precisa criar links de recursos para os bancos de dados e tabelas compartilhados.

1. Na conta de destino (a conta que recebe os recursos compartilhados), crie um link de recurso de banco de dados. Para obter instruções detalhadas, consulte Como criar um link de recurso para um banco de dados compartilhado do catálogo de dados.

   Exemplo de CLI para criar um link de recurso de banco de dados

   ```
   aws glue create-database
   --region us-east-1
   --catalog-id "111122223333"
   --database-input \
   '{
     "Name": "s3table_resourcelink",
     "TargetDatabase": {
       "CatalogId": "011426214932:s3tablescatalog/chmni-s3-table-bucket-011426214932",
       "DatabaseName": "s3_table_ns"
     },
     "CreateTableDefaultPermissions": []
   }'
   ```

2. Conceda permissão entre contas na tabela.

   Exemplo de CLI para concessão de permissão entre contas

   ```
   aws lakeformation grant-permissions \
   --region us-east-1 \
   --cli-input-json \
   '{
       "Principal": {
           "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:role/
   S3TablesTestExecRole"
       },
       "Resource": {
   ```

```
        "Table": {
            "CatalogId": "011426214932:s3tablescatalog/chmni-s3-table-
bucket-011426214932",
            "DatabaseName": "s3_table_ns",
            "Name": "test_s3_iceberg_table"
        }
    },
    "Permissions": [
        "ALL"
    ]
}'
```

3. Conceda a permissão DESCRIBE do Lake Formation no link de recurso.

   Exemplo de CLI para conceder permissão describe no link de recurso.

```
aws lakeformation grant-permissions \
    --principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:role/
S3TablesTestExecRole
    --resource Database='{CatalogId=111122223333;, Name=s3table_resourcelink}' \
    --permissions DESCRIBE
```

Alterando os controles de acesso para integração com tabelas do S3

Depois de integrar as tabelas do Amazon S3 com o AWS Glue Data Catalog, você pode alterar a forma como o acesso aos recursos do seu catálogo é controlado. Esta seção mostra como alterar o controle de acesso dependendo do modelo de controle de acesso atual e desejado. A ativação do Lake Formation permite que você use permissões refinadas, como segurança em nível de coluna e linha, por meio de subsídios do Lake Formation, e permite que o Lake Formation forneça credenciais temporárias em nome dos diretores por meio de uma função registrada. Alterar o controle de acesso de AWS Lake Formation para IAM retorna o controle de acesso às políticas padrão do IAM, o que pode ser apropriado se suas cargas de trabalho não exigirem acesso refinado e você preferir gerenciar as permissões inteiramente por meio do IAM. Ambos os caminhos de migração envolvem a atualização dos padrões do Catálogo de Dados, o ajuste dos registros de recursos com o Lake Formation e a coordenação de concessões de permissões para evitar interrupções no acesso durante a transição.

Tópicos

- Habilite a integração do Lake Formation com o S3 Tables com o Data Catalog
- Alterar o controle de acesso de AWS Lake Formation para o IAM

Habilite a integração do Lake Formation com o S3 Tables com o Data Catalog

Esta seção descreve o fluxo de trabalho para migrar o controle de acesso dos privilégios do IAM para o IAM com AWS Lake Formation concessões para tabelas do Amazon S3 integradas ao. AWS Glue Data Catalog

> **Important**
>
> A ativação do controle de AWS Lake Formation acesso revogará todo o IAM-based acesso existente aos recursos do S3 Tables. Depois de concluir a Etapa 1, os usuários e funções que acessaram dados anteriormente por meio de permissões do IAM perderão o acesso imediatamente. Você deve conceder permissões ao Lake Formation na Etapa 2 antes que os usuários possam consultar os dados novamente. Planeje essa migração durante uma janela de manutenção e coordene com sua equipe de dados.

Pré-requisitos

Para read/write acessar as tabelas do S3, além das permissões do Lake Formation, os diretores também precisam da permissão do `lakeformation:GetDataAccess` IAM. Com essa permissão, o Lake Formation concede a solicitação de credenciais temporárias para acessar os dados.

Utilizar AWS CLI

1. Etapa 1: registrar o bucket no Lake Formation usando a função IAM

   Registre o recurso S3 Tables no Lake Formation.

   > **Note**
   >
   > Se você já tem uma função, certifique-se de que o acesso híbrido seja falso.

   ```
   aws lakeformation register-resource \
   ```

```
  --resource-arn "arn:aws:s3tables:us-east-1:AWSAccountID:bucket/*" \
  --role-arn "arn:aws:iam::AWSAccountID:role/service-role/LFAccessRole" \
  --with-federation
```

2. Etapa 2: Atualizar o AWS Glue catálogo para ativar o controle de acesso do Lake Formation

   Atualize o catálogo com o `CreateDatabaseDefaultPermissions` campo vazio e `CreateTableDefaultPermissions` (definido como`[]`). Isso garante que os bancos de dados e tabelas no catálogo sejam gerenciados usando concessões do Lake Formation em vez das permissões IAM-based padrão.

```
aws glue update-catalog \
  --catalog-id "s3tablescatalog" \
  --catalog-input '{
    "FederatedCatalog": {
        "Identifier": "arn:aws:s3tables:us-east-1:AWSAccountID:bucket/*",
        "ConnectionName": "aws:s3tables"
    },
    "CreateDatabaseDefaultPermissions": [],
    "CreateTableDefaultPermissions": [],
    "AllowFullTableExternalDataAccess": "True"
  }'
```

3. Etapa 3: conceder permissões do Lake Formation à sua equipe de dados

   Conceda permissões do Lake Formation aos diretores (funções, usuários ou grupos) que precisam de acesso. Por exemplo, para conceder acesso de leitura de tabela completa a uma função:

```
aws lakeformation grant-permissions \
  --principal
 DataLakePrincipalIdentifier=arn:aws:iam::AWSAccountID:role/DataTeamRole \
  --resource '{
    "Table": {
        "CatalogId": "AWSAccountID",
        "DatabaseName": "s3tablescatalog/table-bucket-name/namespace",
        "TableWildcard": {}
    }
  }' \
  --permissions "SELECT" "DESCRIBE"
```

   Repita o procedimento para cada combinação principal e de recursos, conforme necessário.

Alterar o controle de acesso de AWS Lake Formation para o IAM

Esta seção descreve o fluxo de trabalho para alterar o controle de acesso de AWS Lake Formation concessões para privilégios do IAM para tabelas do Amazon S3 integradas ao. AWS Glue Data Catalog

> **Important**
>
> Alterar o controle de acesso das AWS Lake Formation concessões para o IAM revogará todo o Formation-based acesso existente do Lake aos seus recursos do S3 Tables. Depois de concluir a Etapa 2, os usuários e funções que acessaram dados anteriormente por meio de subsídios do Lake Formation perderão imediatamente o acesso. Você deve conceder acesso ao IAM na Etapa 1 antes de atualizar o catálogo. Planeje essa migração durante uma janela de manutenção e coordene com sua equipe de dados.

> **Important**
>
> Fine-grained controles de acesso, como acesso em nível de coluna e filtros de células de dados, com objetos do Catálogo de Dados estão disponíveis somente quando usados AWS Lake Formation . Antes de prosseguir com a migração dos controles de acesso do IAM AWS Lake Formation para o IAM, audite suas concessões existentes do Lake Formation usando `aws lakeformation list-permissions` e determine se políticas equivalentes do IAM podem fornecer o acesso de que seus usuários precisam. Qualquer diretor que dependesse de subsídios refinados do Lake Formation exigirá acesso completo ao IAM em nível de tabela após migrar o controle de acesso.

Pré-requisitos

Antes de começar, certifique-se do seguinte:

- Você identificou todos os subsídios do Lake Formation atualmente em vigor para os recursos que estão sendo migrados. Corra `aws lakeformation list-permissions --resource-type TABLE` para revisá-los.
- Você preparou políticas do IAM que fornecem acesso equivalente para todos os principais afetados.

- A função do IAM registrada no Lake Formation ainda existe `lakeformation:GetDataAccess` (necessária durante o período de transição híbrida).

Utilizar AWS CLI

1. Etapa 1: conceder permissões do IAM aos diretores

   Anexe políticas do IAM aos usuários ou funções que precisam de acesso. A política deve incluir permissões de AWS Glue metadados e permissões de dados de tabelas do S3.

   > **Note**
   >
   > O exemplo de política a seguir fornece apenas acesso de leitura.

```
aws iam put-user-policy \
  --user-name GlueIAMAccessUser \
  --policy-name S3TablesIAMAccessPolicy \
  --policy-document '{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "GlueMetadataAccess",
            "Effect": "Allow",
            "Action": [
                "glue:GetCatalog",
                "glue:GetDatabase",
                "glue:GetTable"
            ],
            "Resource": [
                "arn:aws:glue:us-east-1:AWSAccountID:catalog/s3tablescatalog",
                "arn:aws:glue:us-east-1:AWSAccountID:database/
s3tablescatalog/table-bucket-name/namespace",
                "arn:aws:glue:us-east-1:AWSAccountID:table/s3tablescatalog/table-
bucket-name/namespace/*"
            ]
        },
        {
            "Sid": "S3TablesDataAccess",
            "Effect": "Allow",
            "Action": [
```

```
                "s3tables:GetTableBucket",
                "s3tables:GetTable",
                "s3tables:GetTableMetadataLocation",
                "s3tables:GetTableData"
            ],
            "Resource": [
                "arn:aws:s3tables:us-east-1:AWSAccountID:bucket/table-bucket-name",
                "arn:aws:s3tables:us-east-1:AWSAccountID:bucket/table-bucket-name/
table/*"
            ]
        }
    ]
  }'
```

Verifique se todos os usuários e funções afetados podem acessar as tabelas esperadas usando suas credenciais do IAM antes de continuar.

2. Etapa 2: atualizar o catálogo para restaurar as permissões padrão do IAM

   Atualize o catálogo para isso `CreateDatabaseDefaultPermissions` e `CreateTableDefaultPermissions` `ALL` conceda `IAM_ALLOWED_PRINCIPALS` a. `OverwriteChildResourcePermissionsWithDefault`Defina para `Accept` que a alteração se propague para todos os recursos secundários existentes, não apenas para os recém-criados.

```
aws glue update-catalog \
  --catalog-id "s3tablescatalog" \
  --catalog-input '{
    "FederatedCatalog": {
        "Identifier": "arn:aws:s3tables:us-east-1:AWSAccountID:bucket/*",
        "ConnectionName": "aws:s3tables"
    },
    "CreateDatabaseDefaultPermissions": [{
        "Principal": {"DataLakePrincipalIdentifier": "IAM_ALLOWED_PRINCIPALS"},
        "Permissions": ["ALL"]
    }],
    "CreateTableDefaultPermissions": [{
        "Principal": {"DataLakePrincipalIdentifier": "IAM_ALLOWED_PRINCIPALS"},
        "Permissions": ["ALL"]
    }],
    "OverwriteChildResourcePermissionsWithDefault": "Accept"
  }'
```

3. Etapa 3: Cancele o registro do recurso do Lake Formation

Depois de confirmar que todo o acesso está funcionando por meio das políticas do IAM e que nenhum diretor depende dos subsídios do Lake Formation, você pode cancelar o registro do recurso no Lake Formation para concluir a migração.

```
aws lakeformation deregister-resource \
  --resource-arn "arn:aws:s3tables:us-east-1:AWSAccountID:bucket/*"
```

> **Note**
>
> Depois de cancelar o registro do recurso, remova `lakeformation:GetDataAccess` do IAM os diretores que não precisam mais dele.

Nenhuma `revoke-permissions` etapa é necessária.

Criação de um catálogo gerenciado do Amazon Redshift no AWS Glue Data Catalog

Talvez você não tenha um cluster produtor do Amazon Redshift ou uma unidade de compartilhamento de dados do Amazon Redshift disponível atualmente, mas queira criar e gerenciar tabelas do Amazon Redshift usando o AWS Glue Data Catalog. Você pode começar criando um catálogo AWS Glue gerenciado usando a `glue:CreateCatalog` API ou o AWS Lake Formation console definindo o tipo de catálogo como `Managed` e `Catalog source` como Redshift. Essa etapa faz o seguinte:

- Cria um catálogo no Data Catalog
- Registra o catálogo como um local de dados do Lake Formation.
- cria um grupo de trabalho gerenciado do Amazon Redshift sem servidor
- Vincula o grupo de trabalho do Amazon Redshift sem servidor e o Data Catalog usando um objeto de unidade de compartilhamento de dados.

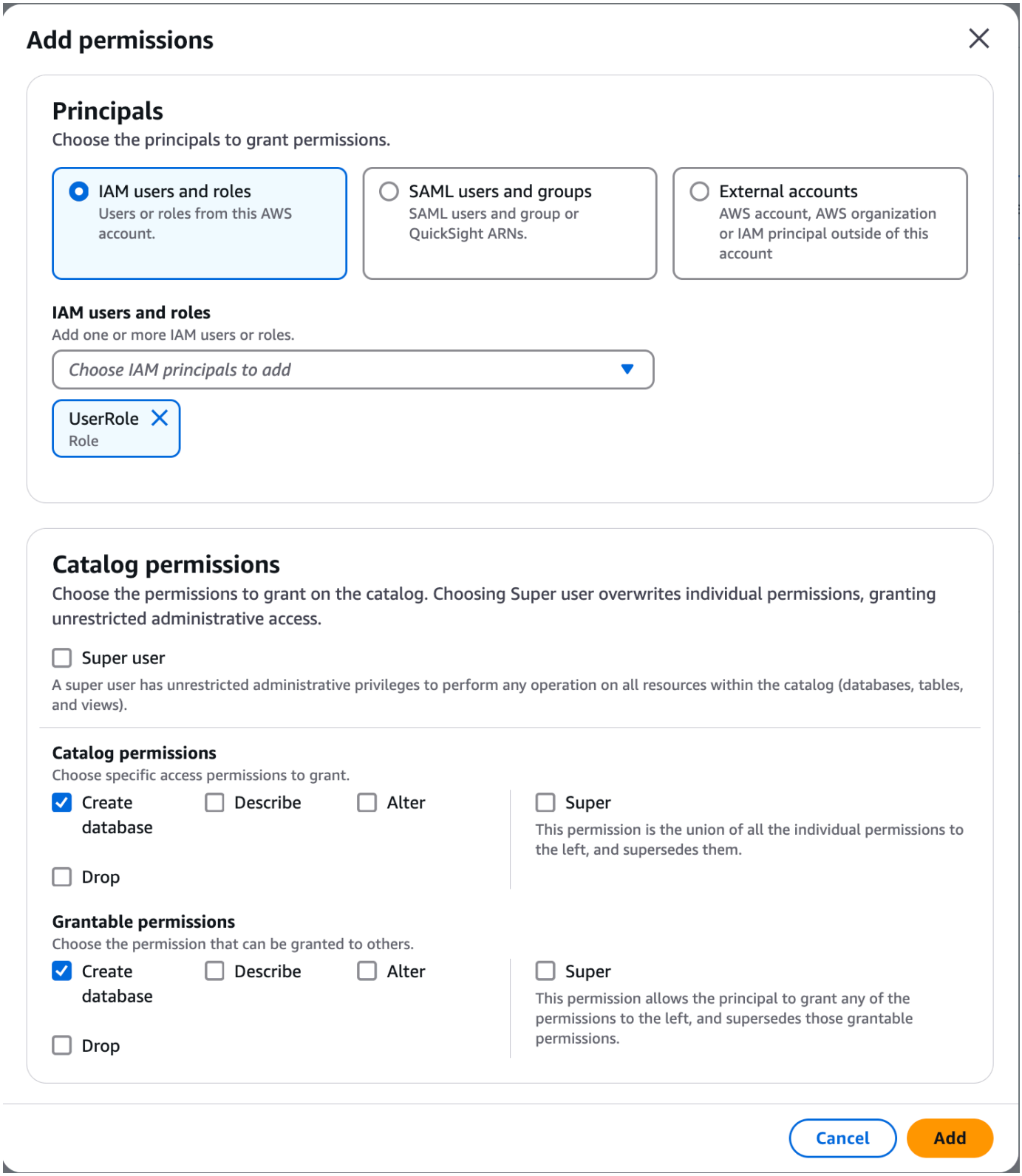
Como criar um catálogo gerenciado e configurar permissões (console)

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.
2. No painel de navegação, escolha Catálogos, em Data Catalog.

3. Selecione a opção Criar catálogo.
4. Na página de detalhes Definir catálogo, insira as seguintes informações:

   - Nome: um nome exclusivo para seu catálogo gerenciado. O nome não pode ser alterado e deve estar em letras minúsculas. O nome pode ter no máximo 255 caracteres.
   - Tipo: escolha `Managed catalog` como o tipo de catálogo.
   - Armazenamento: escolha `Redshift` para armazenamento.
   - Descrição: insira uma descrição para o catálogo criado por meio da fonte de dados.
5. Você pode usar aplicações Apache Spark em execução no Amazon EMR no Amazon EC2 para acessar os bancos de dados Amazon Redshift no AWS Glue Data Catalog.

   Para permitir que o Apache Spark leia e grave no armazenamento gerenciado do Amazon Redshift, o AWS Glue cria um cluster gerenciado do Amazon Redshift com os recursos computacionais e de armazenamento necessários para realizar operações de leitura e gravação sem prejudicar as workloads do data warehouse do Amazon Redshift. Também é preciso fornecer um perfil do IAM com as permissões necessárias para transferir dados de e para o bucket do Amazon S3. Para conhecer as permissões necessárias para o perfil de transferência de dados, consulte a etapa 5 na seção Pré-requisitos para gerenciar namespaces do Amazon Redshift no AWS Glue Data Catalog.
6. Por padrão, os dados no cluster do Amazon Redshift são criptografados usando uma chave AWS gerenciada. O Lake Formation oferece uma opção para criar sua chave do KMS personalizada para criptografia. Se você estiver usando uma chave gerenciada pelo cliente, deverá adicionar políticas de chave específicas à chave.
7. Escolha Personalizar configurações de criptografia se você estiver usando uma chave gerenciada pelo cliente para criptografar os dados no armazenamento gerenciado do Amazon Redshift. Para usar uma chave personalizada, você deve adicionar outra política de chave gerenciada personalizada à sua chave do KMS. Para obter mais informações, consulte Pré-requisitos para gerenciar namespaces do Amazon Redshift no AWS Glue Data Catalog.
8. Opções de criptografia: escolha a opção Personalizar configurações de criptografia se quiser usar uma chave personalizada para criptografar o catálogo. Para usar uma chave personalizada, você deve adicionar outra política de chave gerenciada personalizada à sua chave do KMS.
9. Escolha Próximo para conceder permissões a outras entidades principais.
10. Na página Conceder permissões, escolha Adicionar permissões.
11. Na tela Adicionar permissões, escolha as entidades principais e os tipos de permissões a serem concedidas.



- Na seção Entidades principais, escolha um tipo de entidade principal e, em seguida, especifique as entidades principais para conceder permissões.

- Usuários e perfis do IAM: escolha um ou mais usuários ou perfis na lista de usuários e perfis do IAM.
- Usuários e grupos do SAML — Para SAML e Amazon Quick usuários e grupos, insira um ou mais Amazon Resource Names (ARNs) para usuários ou grupos federados por meio do SAML, ou ARNs para usuários ou grupos do Amazon Quick. Pressione Enter após cada ARN.

  Para obter informações sobre como construir os ARNs, consulte os comandos AWS CLI grant and revoke AWS CLI .

- Na seção Permissões, selecione permissões e permissões concedidas.

  Em Permissões do catálogo, selecione uma ou mais permissões a serem concedidas.

  Escolha Superusuário para conceder permissões administrativas irrestritas em todos os recursos do catálogo.

  Em Permissões concedidas, selecione as permissões que o beneficiário do subsídio pode conceder a outros diretores em sua conta. AWS Essa opção não é compatível quando você está concedendo permissões a uma entidade principal do IAM a partir de uma conta externa.

12. Escolha Próximo para revisar as informações e criar o catálogo. A lista Catálogos mostra o novo catálogo gerenciado.

Como criar um catálogo federado (CLI)

- O exemplo a seguir mostra como criar um catálogo federado.

```
aws glue create-catalog --cli-input-json file://input.json

{
    "Name": "CatalogName",
    "CatalogInput": {
      "Description": "Redshift published Catalog",
      "CreateDatabaseDefaultPermissions" : [],
      "CreateTableDefaultPermissions": [],
      "CatalogProperties": {
        "DataLakeAccessProperties" : {
          "DataLakeAccess" : "true",
          "DataTransferRole" : "DTR arn",
          "KMSKey": "kms key arn",  // Optional
```

```
          "CatalogType": "aws:redshift"
        }
      }
    }
}
```

Resposta get-catalog do Glue

```
aws glue get-catalog \
  --catalog-id account-id:catalog-name \
  --region us-east-1

Response:
{
    "Catalog": {
        "Name": "CatalogName",
        "Description": "Glue Catalog for Redshift z-etl use case",
        "CreateDatabaseDefaultPermissions" : [],
        "CreateTableDefaultPermissions": [],
         "CatalogProperties": {
          "DataLakeAccessProperties" : {
            "DataLakeAccess": "true",
            "DataTransferRole": "DTR arn",
            "KMSKey": "kms key arn",
            "ManagedWorkgroupName": "MWG name",
            "ManagedWorkgroupStatus": "MWG status",
            "RedshiftDatabaseName": "RS db name",
            "NamespaceArn": "namespace key arn",
            "CatalogType": "aws:redshift"
         }
       }
    }
```

Gerenciamento de permissões para dados em uma unidade de compartilhamento de dados do Amazon Redshift.

Com o AWS Lake Formation, você pode gerenciar dados com segurança em uma unidade de compartilhamento de dados do Amazon Redshift. O Amazon Redshift é um serviço de data warehouse totalmente gerenciado e em escala de petabytes na Nuvem AWS. Ao usar o recurso

de compartilhamento de dados, o Amazon Redshift ajuda você a compartilhar dados entre Contas da AWS. Para obter mais informações sobre o compartilhamento de dados do Amazon Redshift, consulte Visão geral do compartilhamento de dados no Amazon Redshift.

No Amazon Redshift, o administrador do cluster produtor cria uma unidade de compartilhamento de dados e a compartilha com o administrador do data lake. Para obter instruções passo a passo sobre como criar um administrador de data lake, consulte Crie um administrador de data lake.

Depois que você (administrador do data lake) aceitar a unidade compartilhamento de dados, deverá criar um banco de dados do AWS Glue Data Catalog para a unidade de compartilhamento de dados específica. Isso serve para que você possa controlar o acesso a ele usando as permissões do Lake Formation. O Lake Formation mapeia cada unidade de compartilhamento de dados para um banco de dados correspondente do catálogo de dados. Estes aparecem como bancos de dados federados no catálogo de dados.

Um banco de dados é chamado de banco de dados federado quando se refere a uma entidade fora do Catálogo de Dados. As tabelas e exibições na unidade de compartilhamento de dados do Amazon Redshift são listadas como tabelas individuais no catálogo de dados. É possível compartilhar o banco de dados federado com entidades principais do IAM e usuários SAML selecionados na mesma conta ou em outra conta com Lake Formation. Você também pode incluir expressões de filtro de linha e coluna para restringir o acesso a determinados dados. Para obter mais informações, consulte Filtragem de dados e segurança por célula no Lake Formation.

Para fornecer acesso aos usuários a uma unidade de compartilhamento de dados do Amazon Redshift, você deve fazer o seguinte:

1. Atualize as Configurações do catálogo de dados para ativar as permissões do Lake Formation.
2. Aceite o convite da unidade de compartilhamento de dados do administrador do cluster produtor do Amazon Redshift e registre a unidade de compartilhamento de dados no Lake Formation.

   Depois de concluir essa etapa, você pode gerenciar a unidade de compartilhamento de dados dentro do Catálogo de Dados do Lake Formation.
3. Crie um banco de dados federado e defina permissões neste banco de dados.
4. Conceda permissões a usuários aos bancos de dados e tabelas. Você pode compartilhar o banco de dados inteiro ou um subconjunto de tabelas com os usuários na mesma conta ou em outra conta.

Para conhecer as limitações, consulte Limitações do compartilhamento de dados do Amazon Redshift.

Tópicos

- Requisitos para configurar permissões em unidades de compartilhamento de dados do Amazon Redshift
- Configuração de permissões para unidades de compartilhamento de dados do Amazon Redshift
- Consultar bancos de dados federados

Requisitos para configurar permissões em unidades de compartilhamento de dados do Amazon Redshift

Atualizar as configurações padrão do catálogo de dados

Para habilitar as permissões do Lake Formation para os recursos do catálogo de dados, recomendamos que você desative as Configurações-padrão do catálogo de dados no Lake Formation. Para obter mais informações, consulte Alterar o modelo de permissão padrão ou usar o modo de acesso híbrido.

Permissões atualizadas

Além das permissões de administrador do data lake (`AWSLakeFormationDataAdmin`), as seguintes permissões também são necessárias para aceitar uma unidade de compartilhamento de dados do Amazon Redshift no Lake Formation:

- `glue:PassConnection on aws:redshift`
- `redshift:AssociateDataShareConsumer`
- `redshift:DescribeDataSharesForConsumer`
- `redshift:DescribeDataShares`

O usuário do IAM administrador do data lake tem as seguintes permissões implicitamente.

- Acesso à localização de dados
- Criar banco de dados
- lakefomation:registerResource

Configuração de permissões para unidades de compartilhamento de dados do Amazon Redshift

Este tópico descreve as etapas que você precisa seguir para aceitar um convite de unidade de compartilhamento de dados, criar um banco de dados federado, e conceder permissões. Você pode usar o console Lake Formation ou o AWS Command Line Interface (AWS CLI). Os exemplos neste tópico mostram o cluster produtor, o catálogo de dados, e o consumidor de dados na mesma conta.

Para saber mais sobre os recursos de entre contas do Lake Formation, consulte Cross-account compartilhamento de dados em Lake Formation.

Como configurar permissões para uma unidade de compartilhamento de dados

1. Analise e aceite um convite de unidade de compartilhamento de dados.

   Console

   1. Faça login no console do Lake Formation como administrador do data lake em https://console.aws.amazon.com/lakeformation/. Navegue até a página Compartilhamento de dados.
   2. Revise as unidades de compartilhamento de dados que você tem autorização para acessar. A coluna Status indica seu status atual de participação na unidade de compartilhamento de dados. O status Pendente indica que você foi adicionado a uma unidade de compartilhamento de dados, mas ainda não o aceitou ou rejeitou o convite.
   3. Para responder ao convite da unidade de compartilhamento de dados, selecione o nome da unidade de compartilhamento de dados e escolha Examinar convite. Em Aceitar ou rejeitar unidade de compartilhamento de dados, examine os detalhes do convite. Selecione Aceitar para aceitar o convite ou Rejeitar para recusar o convite. Se você rejeitar o convite, não terá acesso à unidade de compartilhamento de dados.

   AWS CLI

   Os exemplos a seguir mostram como exibir, aceitar e registrar o convite. Substitua o ID da Conta da AWS por um ID da Conta da AWS válido. Substitua `data-share-arn` pelo nome do recurso da Amazon (ARN) real que faz referência à unidade de compartilhamento de dados.

   1. Visualize um convite pendente.

```
aws redshift describe-data-shares \
 --data-share-arn 'arn:aws:redshift:us-
east-1:111122223333:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
federatedds' \
```

2. Aceitar uma unidade de compartilhamento de dados.

```
 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. Registre a unidade de compartilhamento de dados na conta do Lake Formation. Use a operação da API de RegisterResource para registrar a unidade de compartilhamento de dados no Lake Formation. `DataShareArn` é o parâmetro de entrada para `ResourceArn`.

> **Note**
>
> Este é uma etapa obrigatória.

```
aws lakeformation register-resource \
 --resource-arn 'arn:aws:redshift:us-
east-1:111122223333:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
federatedds'
```

2. Crie um banco de dados.

   Depois de aceitar um convite de unidade de compartilhamento de dados, você precisa criar um banco de dados que aponte para o banco de dados do Amazon Redshift associado à unidade de compartilhamento de dados. Você deve ser administrador de data lake para criar um banco de dados.

   Console

   1. Selecione a unidade de compartilhamento de dados no painel Convites e selecione Definir informações do banco de dados.

2. Em Definir informações do banco de dados, insira um nome e um identificador exclusivos para a unidade de compartilhamento de dados. Você usa esse identificador para mapear a unidade de compartilhamento de dados internamente na hierarquia de metadados (dbName.schema.table).
3. Selecione Avançar para conceder permissões a outros usuários no banco de dados e nas tabelas compartilhados.

AWS CLI

Use o código de exemplo a seguir para criar um banco de dados que aponta para o banco de dados do Amazon Redshift compartilhado com o Lake Formation por meio da 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. Conceder permissões

Depois de criar o banco de dados, você pode conceder permissões a outros usuários em sua conta ou a Contas da AWS e organizações externas. Não será possível conceder permissões de gravação de dados (inserir e excluir) e permissão de metadados (alterar, eliminar e criar) no banco de dados federado que está associado a uma unidade de compartilhamento de dados do Amazon Redshift. Para obter mais informações sobre a concessão de permissões, consulte Gerenciando permissões do Lake Formation.

> **Note**
>
> Como administrador do data lake, você só pode exibir tabelas nos bancos de dados federados. Para executar qualquer outra ação, você precisa conceder a si mais permissões nessas tabelas.

Console

1. Na tela Conceder permissões, selecione os usuários aos quais planeja conceder permissões.
2. Selecione Conceder.

AWS CLI

Use os exemplos a seguir para conceder permissões de banco de dados e tabelas usando a 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"
                   ]
 }
```

Consultar bancos de dados federados

Após conceder permissões, os usuários podem fazer login e começar a consultar o banco de dados federado usando o Amazon Redshift. Agora, os usuários podem usar o nome do banco de dados local para referenciar a unidade de compartilhamento de dados do Amazon Redshift em consultas SQL. No Amazon Redshift, a tabela do cliente no esquema público compartilhada por meio da unidade de compartilhamento de dados terá uma tabela correspondente criada como `public.customer` no catálogo de dados.

1. Antes de consultar o banco de dados federado usando o Amazon Redshift, o administrador do cluster cria um banco de dados a partir do banco de dados do catálogo de dados usando o seguinte comando:

```
CREATE DATABASE sharedcustomerdb FROM ARN
 'arn:aws:glue:<region>:111122223333:database/tahoedb' WITH DATA CATALOG SCHEMA
 tahoedb
```

2. O administrador do cluster concede permissões de uso no banco de dados.

```
GRANT USAGE ON DATABASE sharedcustomerdb TO IAM:user;
```

3. Você (usuário federado) agora pode fazer login nas ferramentas SQL para consultar a tabela.

```
Select * from sharedcustomerdb.public.customer limit 10;
```

Para obter mais informações, consulte Consulta no AWS Glue Data Catalog no Guia de gerenciamento do Amazon Redshift.

Gerenciamento de permissões em conjuntos de dados que usam repositórios de dados externos

Com a federação de AWS Glue Data Catalog metadados (federação do catálogo de dados), você pode conectar o catálogo de dados a metastores externos que armazenam metadados para seus dados do Amazon S3 e gerenciar com segurança as permissões de acesso aos dados usando. AWS Lake Formation Você não precisa migrar os metadados do repositório externo para o catálogo de dados.

O Catálogo de Dados oferece um repositório centralizado de metadados que facilita o gerenciamento e a descoberta de dados em sistemas diferentes. Quando sua organização gerencia dados no catálogo de dados, você pode usar AWS Lake Formation para controlar o acesso aos seus conjuntos de dados no Amazon S3.

> **Note**
>
> Atualmente, oferecemos suporte somente à federação de repositórios do Apache Hive (versão 3 e superior).

Para configurar a federação do Catálogo de Dados, fornecemos um aplicativo AWS Serverless Application Model (AWS SAM) chamado GlueDataCatalogFederation-HiveMetastoreno AWS Serverless Application Repository.

A implementação de referência é fornecida GitHub como um projeto de código aberto na AWS Glue Data Catalog Federation - Hive Metastore.

O AWS SAM aplicativo cria e implanta os seguintes recursos que são necessários para conectar o Catálogo de Dados ao metastore do Hive:

- Uma AWS Lambda função — Hospeda a implementação do serviço de federação que se comunica entre o Catálogo de Dados e o metastore Hive. AWS Glue invoca essa função Lambda para recuperar objetos de metadados do metastore Hive.
- Amazon API Gateway — O endpoint de conexão do seu repositório do Hive que atua como um proxy para rotear todas as invocações para a função do Lambda.
- Um perfil do IAM: um perfil com as permissões necessárias para criar a conexão entre o Catálogo de Dados e o repositório do Hive.
- AWS Glue conexão — Um Amazon API Gateway tipo de AWS Glue conexão que armazena o Amazon API Gateway endpoint e uma função do IAM para invocá-lo.

Quando você consulta tabelas, o AWS Glue serviço faz uma chamada de tempo de execução para o metastore do Hive e busca os metadados. A função do Lambda atua como um tradutor entre o Repositório do Hive e o catálogo de dados.

Após estabelecer a conexão, para sincronizar os metadados na repositório do Hive com o catálogo de dados, você precisa criar um banco de dados federado no catálogo de dados usando as informações da conexão do repositório do Hive, e mapear esse banco de dados para o banco de dados do Hive. Um banco de dados é chamado de banco de dados federado quando aponta para uma entidade fora do catálogo de dados.

Você pode aplicar as permissões do Lake Formation usando o controle de acesso baseado em tags e o método de recurso nomeado no banco de dados federado e compartilhá-lo entre várias Contas da AWS unidades organizacionais (OUs). AWS Organizations Você também pode compartilhar o banco de dados federado diretamente com as entidades principais do IAM de outra conta.

Você pode definir permissões refinadas no nível de coluna, linha e célula usando os filtros de dados do Lake Formation nas tabelas externas do Hive. É possível usar o Amazon Athena, o Amazon Redshift ou o Amazon EMR para consultar as tabelas externas do Hive gerenciadas pelo Lake Formation.

Para obter mais informações sobre compartilhamento de dados entre contas e filtragem de dados, consulte:

- Cross-account compartilhamento de dados em Lake Formation
- Filtragem de dados e segurança por célula no Lake Formation

Etapas de alto nível da federação de metadados do catálogo de dados

1. Você cria usuários e perfis do IAM que têm as permissões apropriadas para implantar a aplicação do AWS SAM e criar bancos de dados federados.
2. Você registra o local dos dados do Amazon S3 com o Lake Formation selecionando a opção `Enable Data Catalog federation` para conjuntos de dados que usam um repositório externo do Hive.
3. Você define as configurações do AWS SAM aplicativo (nome da AWS Glue conexão, URL para o metastore do Hive e parâmetros da função Lambda) e implanta o aplicativo. AWS SAM
4. O AWS SAM aplicativo implanta os recursos necessários para conectar o metastore externo do Hive ao Catálogo de Dados.
5. Para aplicar as permissões do Lake Formation no banco de dados e nas tabelas do Hive, você cria um banco de dados no Catálogo de Dados usando as informações de conexão do repositório do Hive e associa esse banco de dados ao banco de dados do Hive.
6. Conceda permissões nos bancos de dados federados às entidades principais da sua conta ou de outra conta.

> **Note**
>
> Você pode conectar o Data Catalog a um repositório externo do Hive, criar bancos de dados federados e executar consultas e scripts do ETL em bancos de dados e tabelas do Hive sem aplicar as permissões do Lake Formation. Para dados de origem no Amazon S3 que não estão registrados no Lake Formation, o acesso é determinado pelas políticas de permissões do IAM para o Amazon AWS Glue S3 e pelas ações.

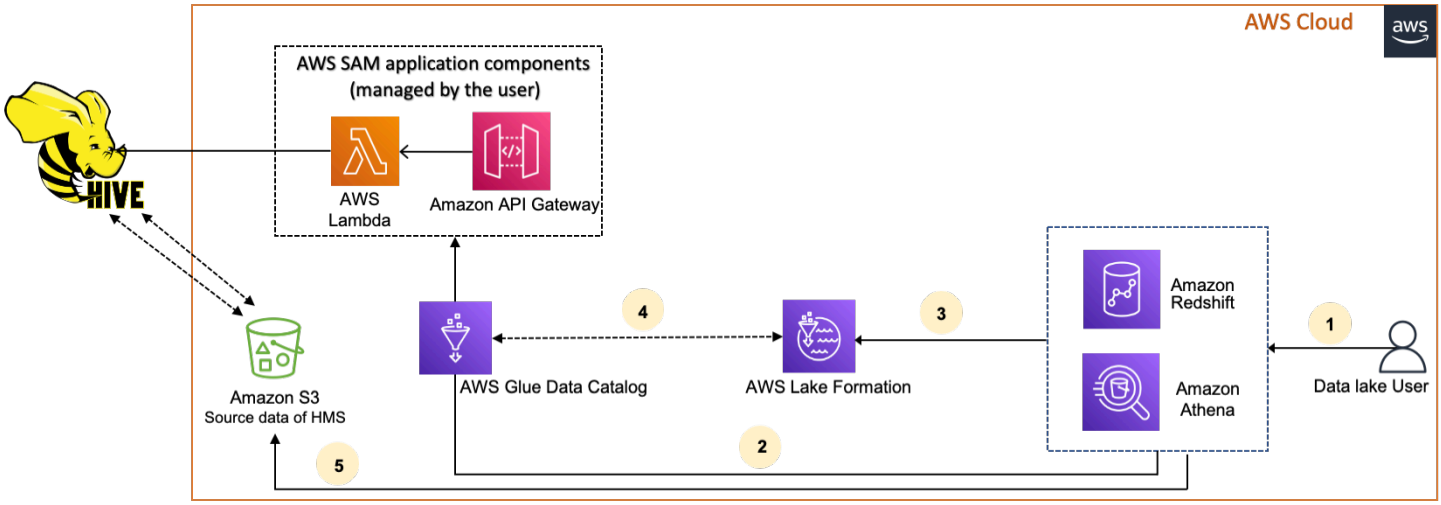
Para conhecer as limitações, consulte Considerações e limitações do compartilhamento de dados de armazenamento de metadados do Hive.

Tópicos

- Fluxo de trabalho
- Requisitos para conectar o catálogo de dados ao Repositório do Hive
- Conectando o catálogo de dados a um repositório externo do Hive
- Recursos adicionais do

Fluxo de trabalho

O diagrama a seguir mostra o fluxo de trabalho para conectar o AWS Glue Data Catalog a um metastore externo do Hive.



1. Uma entidade principal envia uma consulta usando um serviço integrado, como Athena ou Redshift Spectrum.
2. O serviço integrado faz uma chamada para o Catálogo de Dados para obter os metadados, que por sua vez chama o endpoint do metastore Hive disponível por trás Amazon API Gateway e recebe respostas às solicitações de metadados.
3. O serviço integrado envia a solicitação ao Lake Formation para verificar as informações e credenciais da tabela para acessar a tabela.
4. O Lake Formation autoriza a solicitação e fornece credenciais temporárias para o aplicativo integrado, que permite o acesso aos dados.
5. Ao usar as credenciais temporárias recebidas do Lake Formation, o serviço integrado lê os dados do Amazon S3 e compartilha os resultados com a entidade principal.

Requisitos para conectar o catálogo de dados ao Repositório do Hive

Para AWS Glue Data Catalog conectar-se a um metastore externo do Apache Hive e configurar as permissões de acesso aos dados, você precisa preencher os seguintes requisitos:

> **Note**
>
> Recomendamos que um administrador do Lake Formation implante o AWS SAM aplicativo e que somente um usuário privilegiado use a conexão de metastore do Hive para criar os bancos de dados federados correspondentes.

1. Crie perfis do IAM.

   Para implantar o AWS SAM aplicação

   - Crie uma função que tenha as permissões necessárias para implantar os recursos (função Lambda Amazon API Gateway, função do IAM e conexão) necessários para criar uma conexão com AWS Glue a metastore do Hive.

   Como criar bancos de dados federados

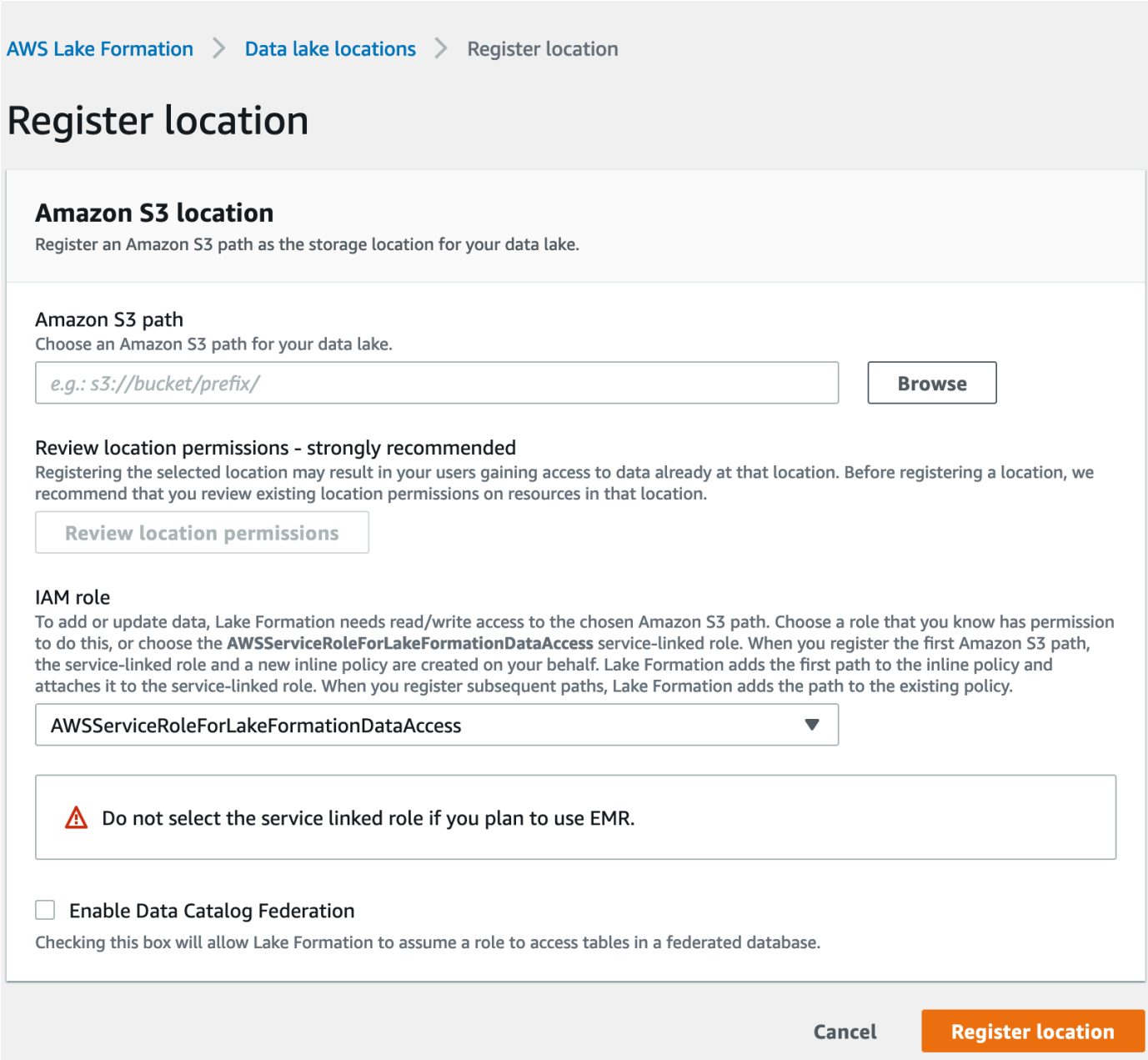
   As seguintes permissões são necessárias nos recursos:

   - `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. Registre o local do Amazon S3 com o Lake Formation.

   Para usar o Lake Formation para gerenciar e proteger os dados em seu data lake, você deve registrar o local do Amazon S3 que tem os dados das tabelas na repositório do Hive com o Lake Formation. Ao fazer isso, a Lake Formation pode fornecer credenciais para serviços AWS analíticos como Athena, Redshift Spectrum e Amazon EMR.

   Para obter mais informações sobre o registro de um local do Amazon S3, consulte Adicionar uma localização do Amazon S3 ao seu data lake.

   Ao registrar a localização do Amazon S3, marque a caixa de seleção Habilitar federação do Catálogo de Dados para permitir que o Lake Formation assuma um perfil para acessar tabelas em um banco de dados federado.



Para obter mais informações sobre como registrar um local de dados no Lake Formation, consulte Como configurar um local no Amazon S3 para o data lake.

3. Use a versão correta do Amazon EMR.

   Para usar o Amazon EMR com os bancos de dados federados do repositório do Hive, você precisa ter o Hive versão 3.x ou posterior e o Amazon EMR versão 6.x ou posterior.

Conectando o catálogo de dados a um repositório externo do Hive

Para conectar o AWS Glue Data Catalog a uma metastore do Hive, você precisa implantar um AWS SAM aplicativo chamado. GlueDataCatalogFederation-HiveMetastore Ele cria os recursos necessários para conectar o repositório externo do Hive ao catálogo de dados. Você pode acessar o AWS SAM aplicativo no AWS Serverless Application Repository.

O AWS SAM aplicativo cria a conexão para o metastore Hive por trás do Amazon API Gateway usando uma função Lambda. O AWS SAM aplicativo usa um identificador uniforme de recursos (URI) como entrada do usuário e conecta o metastore externo do Hive ao Catálogo de Dados. Quando um usuário executa uma consulta nas tabelas do Hive, o Catálogo de Dados chama o endpoint do API Gateway. O endpoint invoca a função do Lambda para recuperar os metadados das tabelas do Hive.

Como conectar o catálogo de dados ao repositório do Hive e configurar permissões

1. Implemente o AWS SAM aplicativo.

    1. Faça login no Console de gerenciamento da AWS e abra AWS Serverless Application Repository o.
    2. No painel de navegação, escolha Aplicativos disponíveis.
    3. Selecione Aplicativos públicos.
    4. Selecione a opção Show apps that create custom IAM roles or resource policies (Mostrar aplicações que criam funções personalizadas do IAM ou políticas de recursos).
    5. Na caixa de pesquisa, insira o nome GlueDataCatalogFederation-HiveMetastore.
    6. Escolha o aplicativo do GlueDataCatalogFederation-HiveMetastore.
    7. Em Configurações do aplicativo, digite as seguintes configurações mínimas necessárias para sua função do Lambda:
        - Nome do aplicativo - Um nome para seu AWS SAM aplicativo.
        - GlueConnectionName- Um nome para a conexão.
        - HiveMetastoreURIs- O URI do seu host de metastore Hive.
        - LambdaMemory- A quantidade de memória Lambda em MB de 128 a 10240. O padrão é 1024.
        - LambdaTimeout- O tempo máximo de execução de invocação do Lambda em segundos. O padrão é 30.
        - VPCSecurityGroupIdse VPCSubnetIds- Informações para o VPC onde existe o metastore Hive.

8. Selecione Reconheço que este aplicativo cria perfis personalizadas do IAM e políticas de recursos. Para obter mais informações, escolha o link Informações.
9. Na seção Configurações do aplicativo, escolha Implantar. Quando a implantação for concluída, a função do Lambda será exibida seção Recursos no console do Lambda.

A aplicação é implantada no Lambda. O nome é prefixado com serverlessrepo- para indicar que a aplicação foi implantada por meio do AWS Serverless Application Repository. Selecionar o aplicativo leva você à página Recursos, na qual cada um dos recursos do aplicativo que foram implantados está listado. Os recursos incluem a função Lambda, que permite a comunicação entre o catálogo de dados e o metastore Hive, a AWS Glue conexão e outros recursos necessários para a federação do banco de dados.

2. Crie um banco de dados federado no catálogo de dados.

Depois de criar uma conexão com o repositório do Hive, você pode criar bancos de dados federados no Catálogo de Dados que apontam para os bancos de dados externos do repositório do Hive. Você precisa criar um banco de dados correspondente no Catálogo de Dados para cada banco de dados de repositório do Hive que você está conectando ao Catálogo de Dados.

Lake Formation console

1. Na página Compartilhamento de dados, selecione a guia Bancos de dados compartilhados e, em seguida, selecione Criar banco de dados.
2. Em Nome da conexão, selecione o nome da sua conexão de repositório do Hive no menu suspenso.
3. Digite um nome de banco de dados exclusivo e o identificador de origem da federação para o banco de dados. Esse é o nome que você usa em suas instruções SQL ao consultar tabelas. O nome pode ter no máximo 255 caracteres e deve ser exclusivo dentro da conta.
4. Selecione Criar banco de dados.

AWS CLI

```
aws glue create-database \
'{
 "CatalogId": "<111122223333>",
  "database-input": {
```

```
    "Name":"<fed_glue_db>",
    "FederatedDatabase":{
        "Identifier":"<hive_db_on_emr>",
        "ConnectionName":"<hms_connection>"
     }
   }
}'
```

3. Visualize tabelas no banco de dados federado.

   Após criar o banco de dados federado, você pode exibir a lista de tabelas em seu repositório do Hive usando o console do Lake Formation ou a AWS CLI.

   Lake Formation console

   1. Selecione o nome do banco de dados na guia Bancos de dados compartilhados.
   2. Na página Bancos de dados, selecione Exibir tabelas.

   AWS CLI

   Os exemplos a seguir mostram como recuperar a definição da conexão, o nome do banco de dados e algumas ou todas as tabelas no banco de dados. Substitua a ID do Catálogo de Dados pela Conta da AWS ID válida que você usou para criar o banco de dados. Substitua `hms_connection` pelo nome da conexão.

```
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. Conceder permissões

   Depois de criar o banco de dados, você pode conceder permissões a outros usuários e funções do IAM em sua conta ou a organizações externas Contas da AWS e externas. Você não poderá conceder permissões de gravação de dados (inserir e excluir) e permissão de metadados (alterar, descartar e criar) nos bancos de dados federados. Para obter mais informações sobre a concessão de permissões, consulte Gerenciando permissões do Lake Formation

5. Consulte os bancos de dados federados.

   Após conceder permissões, os usuários podem fazer login e começar a consultar o banco de dados federado usando o Athena e o Amazon Redshift. Agora, os usuários podem usar o nome do banco de dados local para referenciar o banco de dados do Hive em consultas SQL.

   Exemplo de sintaxe de Amazon Athena consulta

   Substitua o `fed_glue_db` pelo nome do banco de dados local que você criou anteriormente.

   ```
   Select * from fed_glue_db.customers limit 10;
   ```

Recursos adicionais do

A postagem do blog a seguir contém instruções detalhadas para configurar as permissões do Lake Formation em um banco de dados e tabelas de repositório do Hive e consultá-los usando o Athena. Também ilustramos um caso de uso de compartilhamento entre contas, em que um diretor de Lake Formation na conta de produtor A compartilha um banco de dados federado do Hive e tabelas usando LF-Tag a conta de consumidor B.

- Consulte sua metastore do Apache Hive com permissões AWS Lake Formation

Gerenciando permissões do Lake Formation

O Lake Formation fornece controles de acesso central para dados em seu data lake. Você pode definir regras baseadas em políticas de segurança para seus usuários e aplicativos por função no Lake Formation, e a integração com AWS Identity and Access Management autentica esses usuários e funções. Depois que as regras são definidas, o Lake Formation aplica seus controles de acesso em granularidade em nível de tabela, coluna e linha para usuários do Amazon Redshift Spectrum e do Amazon Athena.

Tópicos

- Conceder permissões de localização de dados
- Conceder permissões nos recursos do Catálogo de Dados
- Exemplo de cenário de permissões
- Filtragem de dados e segurança por célula no Lake Formation
- Visualizar permissões de banco de dados e tabelas no Lake Formation
- Revogando a permissão usando o console Lake Formation
- Cross-account compartilhamento de dados em Lake Formation
- Acessar e visualizar tabelas e bancos de dados compartilhados do catálogo de dados
- Criação de links de recursos
- Acessar tabelas entre regiões

Conceder permissões de localização de dados

As permissões de localização de dados AWS Lake Formation permitem que os diretores criem e alterem recursos do catálogo de dados que apontam para locais registrados designados do Amazon S3. As permissões de localização de dados funcionam em conjunto com as permissões de dados do Lake Formation para proteger as informações em seu data lake.

O Lake Formation não usa o serviço AWS Resource Access Manager (AWS RAM) para conceder permissões de localização de dados, então você não precisa aceitar convites de compartilhamento de recursos para obter permissões de localização de dados.

Você pode conceder permissões de localização de dados usando o console do Lake Formation, a API ou AWS Command Line Interface (AWS CLI).

> **Note**
>
> Para que uma concessão seja bem-sucedida, você deve primeiro registrar a localização dos dados no Lake Formation.

> **Consulte também:**
>
> - Underlying data access control

Tópicos

- Concessão de permissões de localização de dados (mesma conta)
- Concessão de permissões de localização de dados (conta externa)
- Conceder permissões em um local de dados compartilhado com sua conta

Concessão de permissões de localização de dados (mesma conta)

Siga estas etapas para conceder permissões de localização de dados às entidades principais da sua conta da AWS . Você pode conceder permissões usando o console do Lake Formation, a API ou a AWS Command Line Interface (AWS CLI).

Console de gerenciamento da AWS

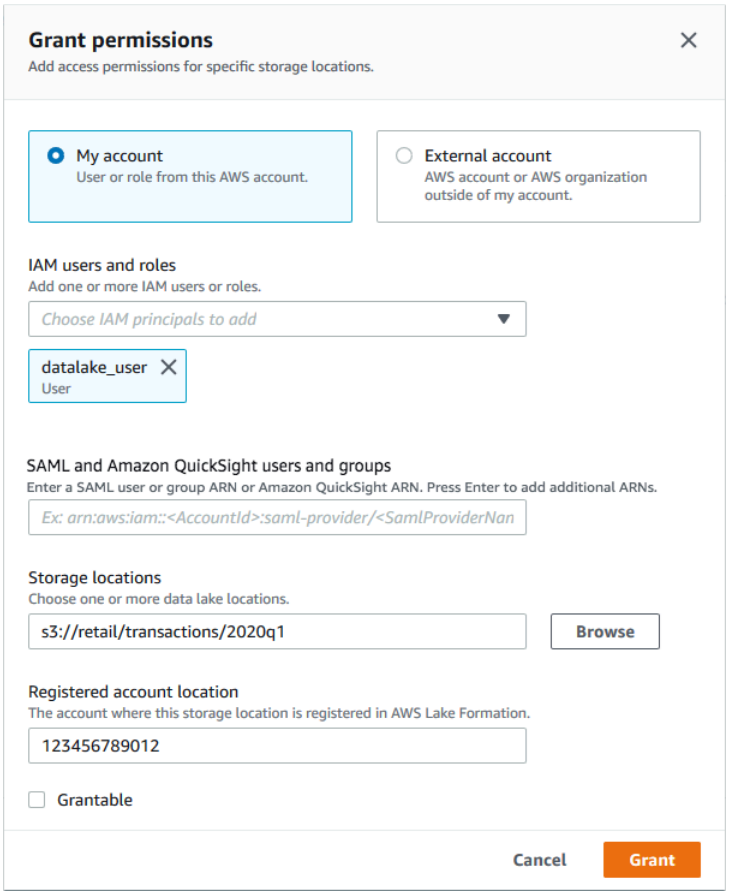
Como conceder permissões de localização de dados (mesma conta)

1. Abra o AWS Lake Formation console em https://console.aws.amazon.com/lakeformation/. Faça login como administrador do data lake ou como entidade principal que concedeu permissões no local de dados desejado.
2. No painel de navegação, em Permissões, selecione Locais de dados.
3. Selecione Conceder.
4. Na caixa de diálogo Conceder permissões, verifique se o quadro Minha conta está selecionado. Em seguida, forneça as seguintes informações:
   - Para usuários e funções do IAM, escolha um ou mais entidades principais.

- Para usuários e grupos do SAML e do Amazon Quick, insira um ou mais nomes de recursos da Amazon (ARNs) para usuários ou grupos federados por meio de SAML ou ARNs para usuários ou grupos do Amazon Quick.

  Insira um ARN por vez e pressione Enter após cada ARN. Para obter informações sobre como construir os ARNs, consulte Concessão e revogação da Lake Formation AWS CLI Comandos da do.

- Para Locais de armazenamento, escolha Browse e escolha um local de armazenamento do Amazon Simple Storage Service (Amazon S3). O local deve ser registrado no Lake Formation. Escolha Procurar novamente para adicionar outro local. Você também pode digitar o local, mas certifique-se de preceder o local com `s3://`.
- Em Local da conta registrada, insira o ID da AWS conta em que o local está registrado. O padrão é o ID da sua conta. Em um cenário de várias contas, os administradores de data lake em uma conta de destinatário podem especificar a conta do proprietário aqui ao conceder a permissão de localização de dados a outras entidades principais na conta do destinatário.
- (Opcional) Para permitir que as entidades principais selecionadas concedam permissões de localização de dados no local selecionado, escolha Concedível.



5. Selecione Conceder.

AWS CLI

Como conceder permissões de localização de dados (mesma conta)

- Execute um comando `grant-permissions` e conceda `DATA_LOCATION_ACCESS` à entidade principal, especificando o caminho do Amazon S3 como o recurso.

  Example

  O exemplo a seguir concede permissões de localização de dados em `s3://retail` ao usuário `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

O exemplo a seguir concede permissões de localização de dados em `s3://retail` ao grupo `ALLIAMPrincipals`.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333:IAMPrincipals --
permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"ResourceArn":"arn:aws:s3:::retail", "CatalogId": "111122223333"}}'
```

Consulte também:

- Referência de permissões do Lake Formation

Concessão de permissões de localização de dados (conta externa)

Siga estas etapas para conceder permissões de localização de dados a uma AWS conta ou organização externa.

Você pode conceder permissões usando o console do Lake Formation, a API ou a AWS Command Line Interface (AWS CLI).

Antes de começar

Certifique-se de que todos os pré-requisitos de acesso entre contas sejam atendidos. Para obter mais informações, consulte Pré-requisitos.

Console de gerenciamento da AWS

Para conceder permissões de localização de dados (conta externa, console)

1. Abra o AWS Lake Formation console em https://console.aws.amazon.com/lakeformation/. Faça login como administrador de data lake.
2. No painel de navegação, em Permissões, escolha Localizações dos dados e, em seguida, escolha Conceder.

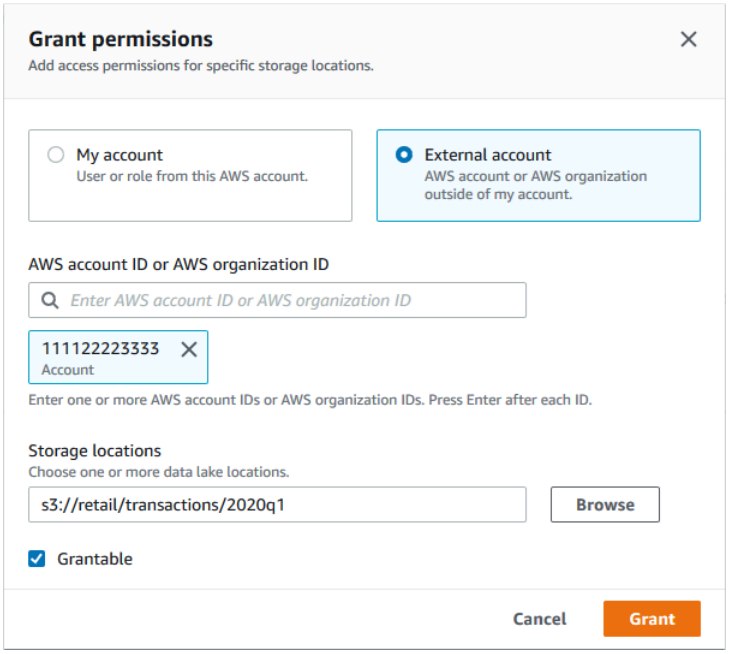
3. Na caixa de diálogo Conceder permissões, escolha o quadro Conta externa.
4. Forneça as informações a seguir:

   - Para ID AWS da conta ou ID AWS da organização, insira números de AWS conta, IDs de organização ou IDs de unidade organizacional válidos.

     Pressione Enter após cada ID.

     O ID da organização consiste em "o-" seguido por 10 a 32 letras minúsculas ou dígitos.

     Um ID de unidade organizacional consiste em "ou-" seguido de 4 a 32 letras minúsculas ou dígitos (o ID da raiz que contém a OU). Essa string é seguida por um segundo "-" (hífen) e 8 a 32 letras minúsculas ou dígitos adicionais.
   - Em Locais de armazenamento, escolha Procurar e escolha um local de armazenamento do Amazon Simple Storage Service (Amazon S3). O local deve ser registrado no Lake Formation.



5. Selecione Concedível.
6. Selecione Conceder.

AWS CLI

Para conceder permissões de localização de dados (conta externa, AWS CLI)

- Para conceder permissões a uma AWS conta externa, digite um comando semelhante ao seguinte.

```
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"}}'
```

Esse comando concede a `DATA_LOCATION_ACCESS` a opção de concessão à conta 1111-2222-3333 no local `s3://retail/transactions/2020q1` do Amazon S3, que pertence à conta 1234-5678-9012.

Para conceder permissões a uma organização, digite um comando semelhante ao seguinte:

```
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"}}'
```

Este comando concede a `DATA_LOCATION_ACCESS` a opção de concessão à organização `o-abcdefghijkl` no local do Amazon S3 `s3://retail/transactions/2020q1`, que pertence à conta 1234-5678-9012.

Para conceder permissões a um principal em uma AWS conta externa, digite um comando semelhante ao seguinte.

```
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"}}'
```

Esse comando é concedido `DATA_LOCATION_ACCESS` a uma entidade principal na conta 1111-2222-3333 na localização `s3://retail/transactions/2020q1` do Amazon S3, que pertence à conta 1234-5678-9012.

Example

O exemplo a seguir concede permissões de localização de dados a `s3://retail` para um grupo `ALLIAMPrincipals` em uma conta externa.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333:IAMPrincipals --
permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"ResourceArn":"arn:aws:s3:::retail", "CatalogId": "123456789012"}}'
```

Consulte também:

- Referência de permissões do Lake Formation

Conceder permissões em um local de dados compartilhado com sua conta

Depois que um recurso do Catálogo de Dados é compartilhado com sua AWS conta, como administrador do data lake, você pode conceder permissões sobre o recurso a outros diretores da sua conta. Se a permissão `ALTER` for concedida em uma tabela compartilhada e a tabela apontar para um local registrado no Amazon S3, você também deverá conceder permissões de localização de dados no local. Da mesma forma, se a permissão `CREATE_TABLE` ou `ALTER` for concedida em um banco de dados compartilhado e o banco de dados tiver uma propriedade de localização que aponte para um local registrado, você também deverá conceder permissões de localização de dados no local.

Para conceder permissões de localização de dados em um local compartilhado a uma entidade principal em sua conta, sua conta deve ter recebido a permissão `DATA_LOCATION_ACCESS` no local com a opção de concessão. Ao conceder `DATA_LOCATION_ACCESS` a outro principal em sua conta, você deve incluir a ID do catálogo de dados (ID da AWS conta) da conta do proprietário. A conta do proprietário é a conta que registrou o local.

Você pode usar o AWS Lake Formation console, a API ou o AWS Command Line Interface ()AWS CLI para conceder permissões de localização de dados.

Para conceder permissões em um local de dados compartilhado com sua conta (console)

- Siga as etapas em Concessão de permissões de localização de dados (mesma conta).

  Para Locais de armazenamento, você deve digitar os locais. Em Localização da conta registrada, insira o AWS ID da conta do proprietário.

Para conceder permissões em um local de dados compartilhado com sua conta (AWS CLI)

- Digite um dos comandos a seguir para conceder permissões a um usuário ou a uma função.

```
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>"}}'
```

Conceder permissões nos recursos do Catálogo de Dados

Você pode conceder permissões de dados aos diretores para que eles AWS Lake Formation possam criar e gerenciar recursos do Catálogo de Dados e acessar os dados subjacentes. É possível conceder Permissões de data lake em catálogos, bancos de dados, tabelas e visualizações. Ao conceder permissões em tabelas, é possível limitar o acesso a colunas ou linhas específicas da tabela para um controle de acesso ainda mais refinado.

É possível conceder permissões em catálogos, bancos de dados, tabelas e visualizações individuais ou, com uma única operação de concessão, você pode conceder permissões em todos os bancos de dados, tabelas e visualizações em um catálogo ou banco de dados. Se você conceder permissões em todas as tabelas em um banco de dados a entidades principais do IAM, estará concedendo implicitamente a permissão `DESCRIBE` no banco de dados. Em seguida, o banco de dados aparece na página Bancos de dados no console e é retornado pela operação da API `GetDatabases`. O

mesmo princípio se aplica em nível de catálogo: quando você recebe permissões para bancos de dados em um catálogo, também recebe permissões `DESCRIBE` para esse catálogo.

> **Important**
>
> A `DESCRIBE` permissão implícita se aplica somente ao conceder permissões aos diretores do IAM na mesma conta. AWS Em relação a recursos entre contas, você deve conceder permissões `DESCRIBE` explicitamente. A concessão da permissão `DESCRIBE` automática não se aplica quando você usa o controle de acesso por atributo (ABAC). Quando você concede permissões em todas as tabelas em um banco de dados usando atributos, o Lake Formation não concede a permissão `DESCRIBE` implícita ao banco de dados.

Você pode conceder permissões usando o método de recurso nomeado ou o método de controle de acesso (LF-TBAC) baseado na tag Lake Formation.

Você pode conceder permissões aos diretores nas mesmas contas Conta da AWS ou organizações externas. Quando você as concede a contas ou organizações externas, você está compartilhando objetos do Data Catalog que possui com essas contas ou organizações. As entidades principais dessas contas ou organizações podem então acessar os objetos do Data Catalog que você possui e os dados subjacentes.

> **Note**
>
> Atualmente, o LF-TBAC método oferece suporte à concessão de permissões entre contas para diretores Contas da AWS, organizações e unidades organizacionais (OUs) do IAM.

Ao conceder permissões a contas ou organizações externas, você deve incluir a opção de concessão. Somente o administrador do data lake na conta externa pode acessar os objetos compartilhados até que o administrador conceda permissões nos objetos compartilhados a outras entidades principais na conta externa.

Você pode conceder permissões do Catálogo de Dados usando o AWS Lake Formation console, a API ou o AWS Command Line Interface (AWS CLI).

> **Note**
>
> Ao excluir um recurso do Data Catalog, todas as permissões associadas ao objeto se tornam inválidas. Recriar o mesmo recurso com o mesmo nome não recuperará as permissões do Lake Formation. Os usuários precisarão configurar novas permissões novamente.

> **Consulte também:**
>
> - Compartilhando tabelas e bancos de dados do Catálogo de Dados entre AWS Contas
> - Controle de acesso a metadados
> - Referência de permissões do Lake Formation

Permissões do IAM necessárias para conceder ou revogar as permissões do Lake Formation

Todos os diretores, incluindo o administrador do data lake, precisam das seguintes permissões AWS Identity and Access Management (IAM) para conceder ou revogar as permissões do catálogo de AWS Lake Formation dados ou as permissões de localização de dados com a API Lake Formation ou com: AWS CLI

- `lakeformation:GrantPermissions`
- `lakeformation:BatchGrantPermissions`
- `lakeformation:RevokePermissions`
- `lakeformation:BatchRevokePermissions`
- `glue:GetTable`, `glue:GetDatabase` ou `glue:GetCatalog` para uma tabela, banco de dados ou catálogo para o qual você está concedendo permissões com o método de recurso nomeado.

> **Note**
>
> Os administradores do data lake possuem permissões implícitas do Lake Formation para conceder e revogar permissões do Lake Formation. Mas eles ainda precisam das permissões do IAM nas operações de concessão e revogação da API do Lake Formation.

As funções do IAM com política `AWSLakeFormationDataAdmin` AWS gerenciada não podem adicionar novos administradores de data lake porque essa política contém uma negação explícita para a operação da API Lake Formation,. `PutDataLakeSetting`

A política do IAM a seguir é recomendada para entidades principais que não são administradores de data lake e que desejam conceder ou revogar permissões usando o console do Lake Formation.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:ListPermissions",
                "lakeformation:GrantPermissions",
                "lakeformation:BatchGrantPermissions",
                "lakeformation:RevokePermissions",
                "lakeformation:BatchRevokePermissions",
                "glue:GetCatalogs",
                "glue:GetDatabases",
                "glue:SearchTables",
                "glue:GetTables",
                "glue:GetCatalog",
                "glue:GetDatabase",
                "glue:GetTable",
                "iam:ListUsers",
                "iam:ListRoles",
                "sso-directory:DescribeUser",
                "sso-directory:DescribeGroup",
                "sso:DescribeInstance"
            ],
            "Resource": "*"
        }
    ]
}
```

Todas `glue:` as `iam:` permissões desta política estão disponíveis na política AWS gerenciada`AWSGlueConsoleFullAccess`.

Para conceder permissões usando o controle de acesso (LF-TBAC) baseado em tags do Lake Formation, os diretores precisam de permissões adicionais do IAM. Para obter mais informações, consulte Considerações e práticas recomendadas de controle de acesso com base em tags do Lake Formation e Referência de personas e permissões do IAM do Lake Formation.

Cross-account permissões

Os usuários que desejam conceder permissões entre contas do Lake Formation usando o método de recurso nomeado também devem ter as permissões na política `AWSLakeFormationCrossAccountManager` AWS gerenciada.

Os administradores do Data Lake precisam dessas mesmas permissões para conceder permissões entre contas, além da permissão AWS Resource Access Manager (AWS RAM) para permitir a concessão de permissões às organizações. Para obter mais informações, consulte Permissões de administrador do data lake.

O usuário administrador

Um diretor com permissões administrativas — por exemplo, com a política `AdministratorAccess` AWS gerenciada — tem permissões para conceder permissões do Lake Formation e criar administradores de data lake. Para negar a um usuário ou função o acesso às operações de administrador do Lake Formation, anexe ou adicione à política uma declaração `Deny` para as operações da API do administrador.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Action": [
                "lakeformation:GetDataLakeSettings",
                "lakeformation:PutDataLakeSettings"
            ],
            "Effect": "Deny",
            "Resource": [
                "*"
```

```
            ]
        }
    ]
}
```

> **Important**
>
> Para evitar que os usuários se adicionem como administradores com um script de extração, transformação e carregamento (ETL), certifique-se de que todos os usuários e funções não administradores tenham acesso negado a essas operações de API. A política `AWSLakeFormationDataAdmin` AWS gerenciada contém uma negação explícita da operação da API Lake Formation, `PutDataLakeSetting` que impede que os usuários adicionem novos administradores de data lake.

Conceder permissões de dados usando o método de recurso nomeado

O método de recurso nomeado do Data Catalog é uma forma de conceder permissões a objetos do AWS Glue Data Catalog , como catálogos, bancos de dados, tabelas, colunas e visualizações, usando uma abordagem centralizada. Ele permite que você defina políticas baseadas em recursos que controlam o acesso a recursos específicos no data lake.

Ao usar o método de recurso nomeado para conceder permissões, você pode especificar o tipo de recurso e as permissões que deseja conceder ou revogar para esse recurso. Você também pode revogar a permissão posteriormente, se necessário, removendo, assim, as permissões dos recursos associados.

Você pode conceder permissões usando o AWS Lake Formation console, as APIs ou o AWS Command Line Interface (AWS CLI).

Tópicos

- Conceder permissões do catálogo usando o método de recurso nomeado
- Conceder permissões de banco de dados usando o método de recurso nomeado
- Conceder permissões de tabela usando o método de recurso nomeado
- Conceder permissões em visualizações usando o método de recurso nomeado

Conceder permissões do catálogo usando o método de recurso nomeado

As etapas a seguir explicam como conceder permissões de catálogo usando o método de recurso nomeado.

Console

Use a página Conceder permissões no console do Lake Formation. A página está dividida nas seguintes seções:

- Tipo de entidade principal: você pode conceder permissões a entidades principais específicas ou usar tags de atributo.
  - Entidades principais: usuários e perfis do IAM, usuários e grupos do Centro de Identidade do IAM, usuários e grupos do SAML, contas da AWS , organizações ou unidades organizacionais aos quais conceder permissões.

    Entidade principal por atributos: adicione pares de chave-valor de tag de IAMroles ou de sessão do IAM. As entidades principais com atributos correspondentes recebem acesso ao recurso especificado.
  - LF-Tags ou recursos do catálogo — Os catálogos, bancos de dados, tabelas, visualizações ou links de recursos nos quais conceder permissões.
  - Permissões – As permissões do Lake Formation devem ser concedidas.

> **Note**
>
> Para conceder permissões em um link de recurso de banco de dados, consulte Como conceder permissões de links de recursos.

1. Abra a página Conceder permissões.

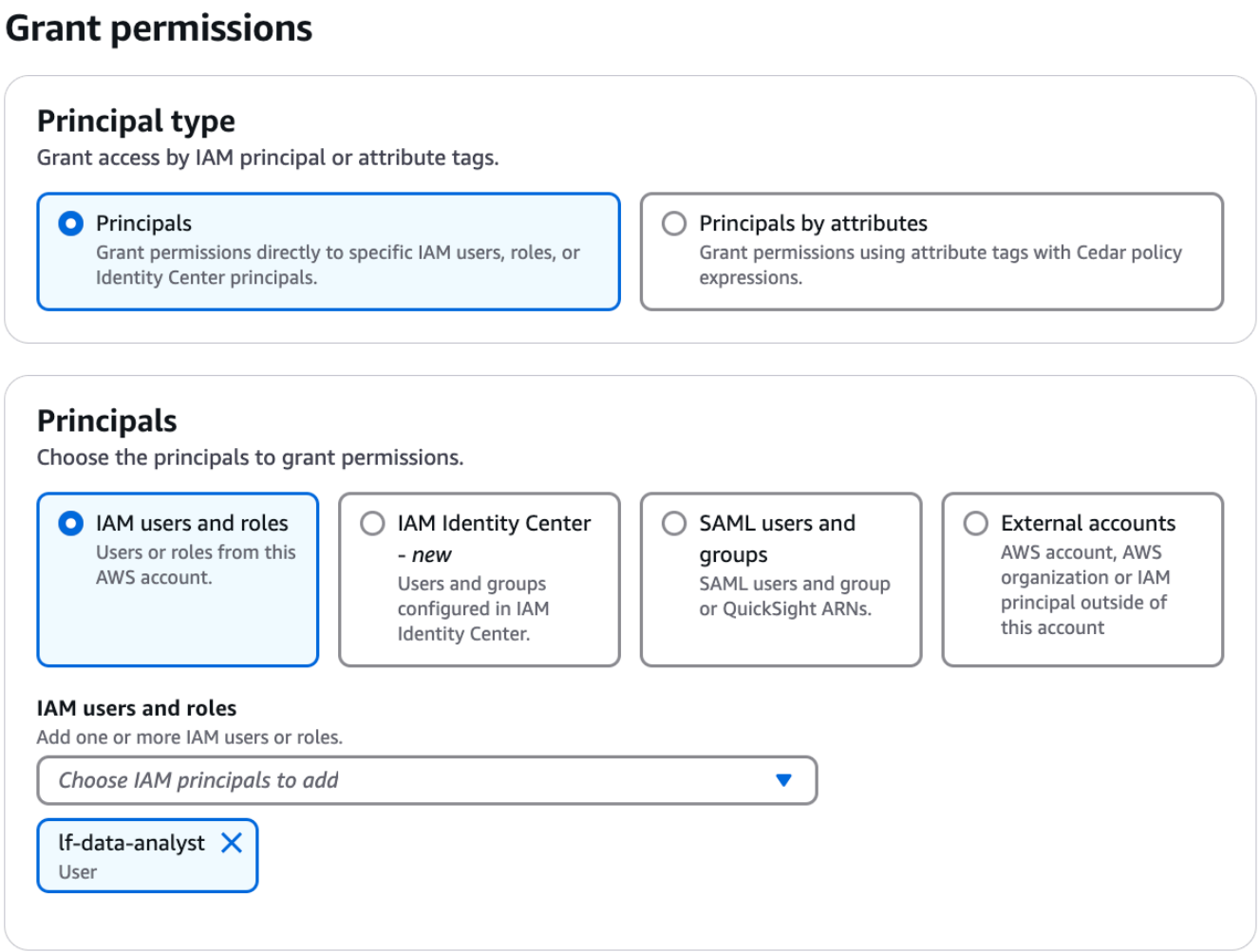
   Abra o AWS Lake Formation console em https://console.aws.amazon.com/lakeformation/e faça login como administrador do data lake, criador do catálogo ou usuário do IAM com permissões concedidas no catálogo.

   Execute um destes procedimentos:

- No painel de navegação, em Permissões, escolha Permissões de dados. Em seguida, escolha Conceder.
- No painel de navegação, escolha Catálogos, em Data Catalog. Depois, na página Catálogos, selecione um catálogo e, no menu Ações, em Permissões, escolha Conceder.

> **Note**
>
> É possível conceder permissões em um catálogo por meio do link do recurso. Para fazer isso, na página Catálogos, escolha um contêiner de links de catálogo e no menu Ações, escolha Conceder no destino. Para obter mais informações, consulte Como os links de recursos funcionam no Lake Formation.

2. Depois, na seção Tipo de entidade principal, escolha as entidades principais ou especifique os atributos anexados a elas.



Especifique os diretores

Usuários e perfis do IAM

Escolha um ou mais usuários ou perfis na lista de usuários e perfis do IAM.

Centro de Identidade do IAM

Selecione um ou mais usuários ou grupos na lista Usuários e grupos. Selecione Adicionar para adicionar mais usuários ou grupos.

Usuários e grupos SAML

Para usuários e grupos do SAML e do Quick, insira um ou mais nomes de recursos da Amazon (ARNs) para usuários ou grupos federados por meio do SAML ou ARNs para usuários ou grupos do Amazon Quick. Pressione Enter após cada ARN.

Para obter informações sobre como construir os ARNs, consulte Concessão e revogação da Lake Formation AWS CLI Comandos da do.

> **Note**
>
> A integração do Lake Formation com o Quick é compatível somente com o Quick Enterprise Edition.

Contas externas

Para Conta da AWS, AWS organização ou diretor do IAM, insira um ou mais IDs de AWS conta, IDs de organização, IDs de unidade organizacional ou ARN válidos para o usuário ou função do IAM. Pressione Enter após cada ID.

O ID da organização consiste em “o-” seguido por 10 a 32 letras minúsculas ou dígitos.

Uma ID de unidade organizacional começa com “ou-” seguida de 4 a 32 letras minúsculas ou dígitos (o ID da raiz que contém a OU). Essa sequência é seguida por um segundo travessão "-" e 8 a 32 letras minúsculas ou dígitos adicionais.

Entidades principais por atributos

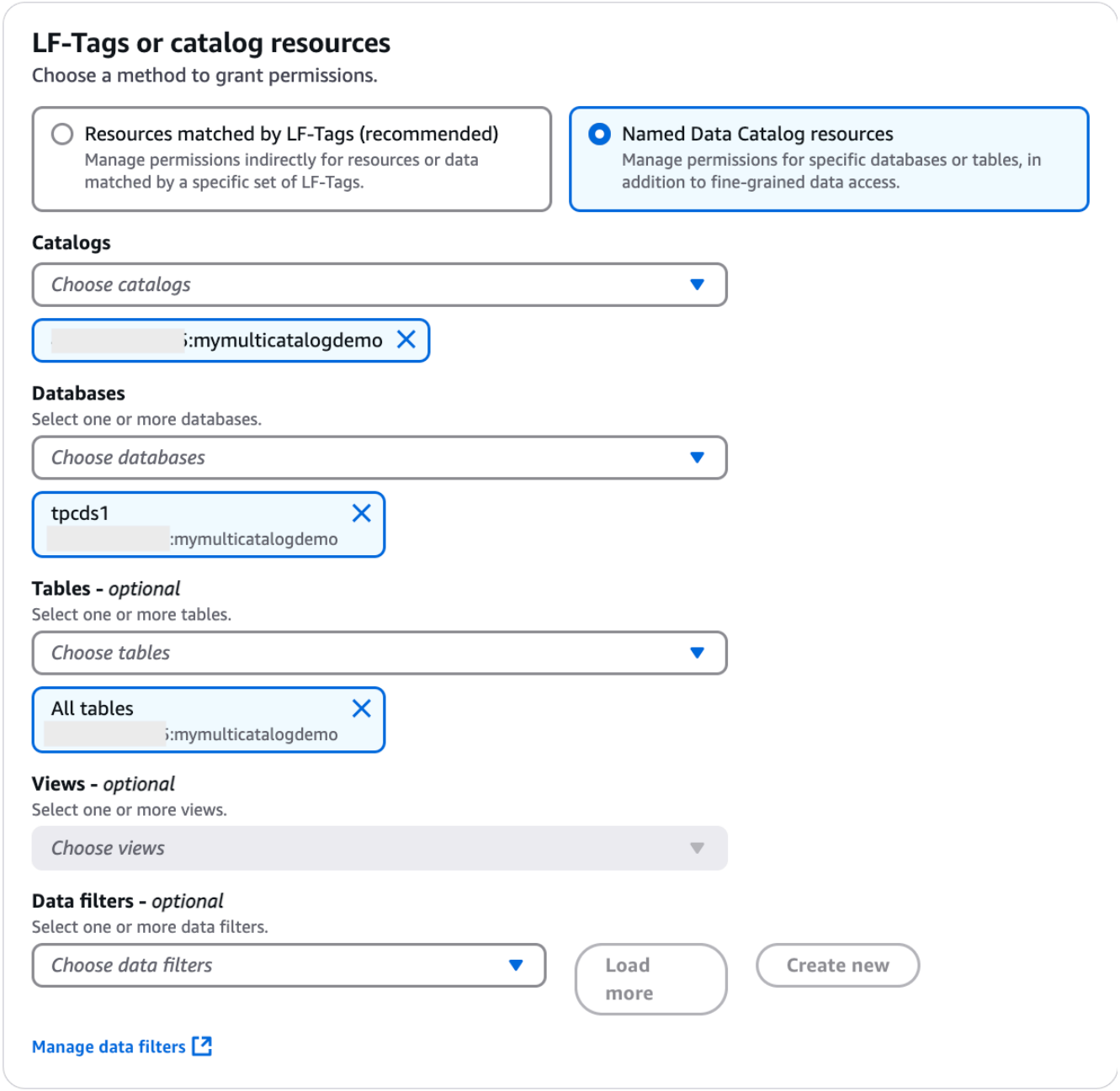
Atributos.

Adicione os pares de chave-valor de tags do IAM por meio do perfil do IAM.

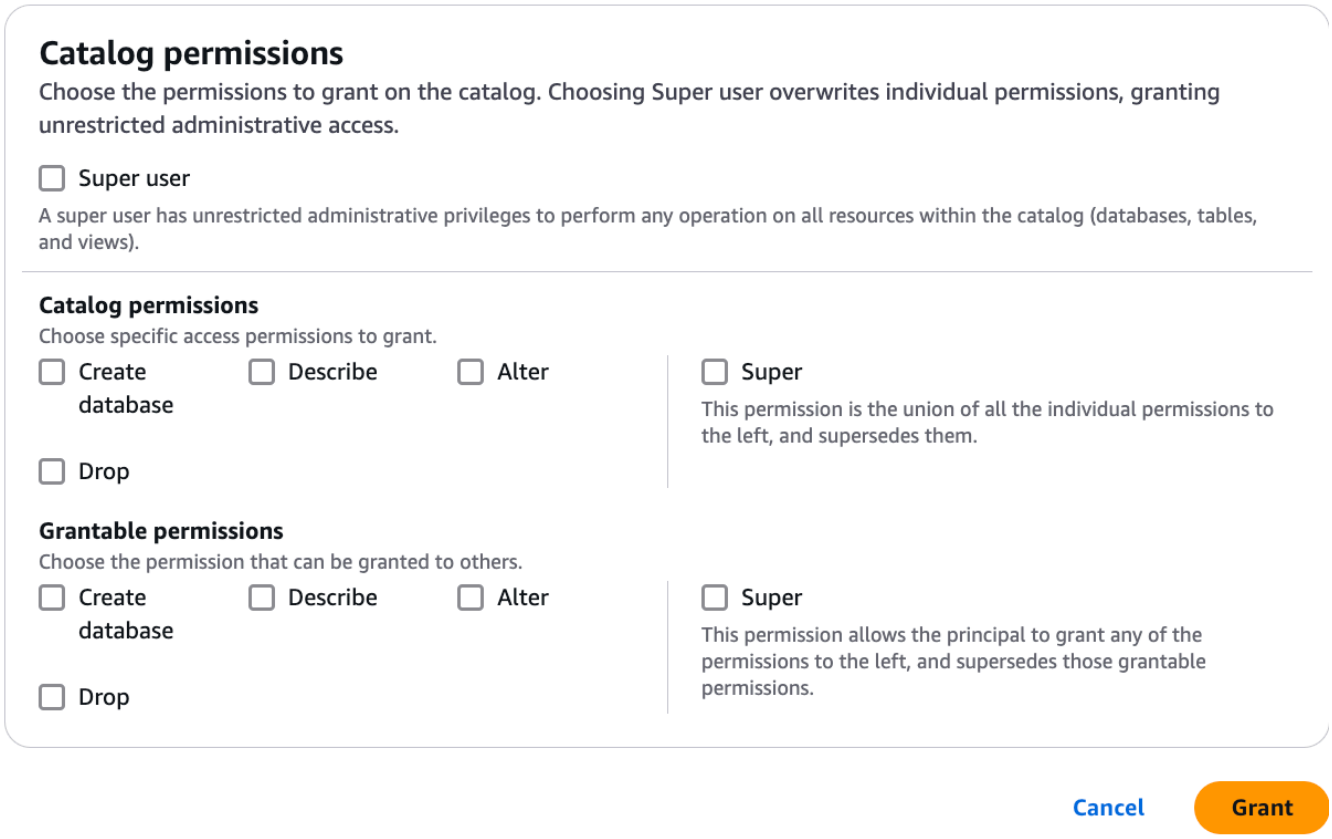
Escopo de permissões

Especifique se você está concedendo permissões a entidades principais com atributos correspondentes na mesma ou em outra conta.

3. Na seção LF-Tags ou recursos do catálogo, escolha Recursos do catálogo de dados nomeados.



4. Escolha um ou mais catálogos na lista Catálogos. Você também pode escolher um ou mais bancos de dados, tabelas e filtros and/or de dados.
5. Na seção Permissões do catálogo, selecione permissões e permissões para concessão. Em Permissões do catálogo, selecione uma ou mais permissões a serem concedidas.



Escolha Superusuário a fim de conceder privilégios administrativos irrestritos para realizar qualquer operação em todos os recursos do catálogo (bancos de dados, tabelas e visualizações).

> **Note**
>
> Depois de conceder `Create database` ou `Alter` em um catálogo que tenha uma propriedade de localização que aponta para um local registrado, certifique-se também de conceder permissões de localização de dados no local às entidades principais. Para obter mais informações, consulte Conceder permissões de localização de dados.

6. (Opcional) Em Permissões concedidas, selecione as permissões que o beneficiário da concessão pode conceder a outras entidades principais em sua conta da AWS . Essa opção não é compatível quando você está concedendo permissões a uma entidade principal do IAM a partir de uma conta externa.
7. Selecione Conceder.

A página Permissões de dados mostra os detalhes da permissão. Se você usou a opção Entidades principais por atributo para conceder permissões, poderá visualizar a permissão concedida a `ALLPrincipals` na lista.

AWS CLI

Para conceder permissões de catálogo usando AWS CLI, consulteCriar catálogos federados do Amazon Redshift.

Conceder permissões de banco de dados usando o método de recurso nomeado

As etapas a seguir explicam como conceder permissões de banco de dados usando o método de recurso nomeado.

Console

Use a página Conceder permissões no console do Lake Formation. A página está dividida nas seguintes seções:

- Tipo de entidade principal: a seção Entidades principais incluem os usuários, os perfis do IAM, usuários e grupos do Centro de Identidade do IAM, usuários e grupos do SAML, contas da AWS , organizações ou unidades organizacionais aos quais conceder permissões. Na seção Entidades principais por atributos, é possível especificar a chave e os valores dos atributos anexados aos perfis do IAM.
- LF-Tags ou recursos do catálogo — Os bancos de dados, tabelas, visualizações ou links de recursos nos quais conceder permissões.
- Permissões – As permissões do Lake Formation devem ser concedidas.

> **Note**
>
> Para conceder permissões em um link de recurso de banco de dados, consulte Como conceder permissões de links de recursos.

1. Abra a página Conceder permissões.

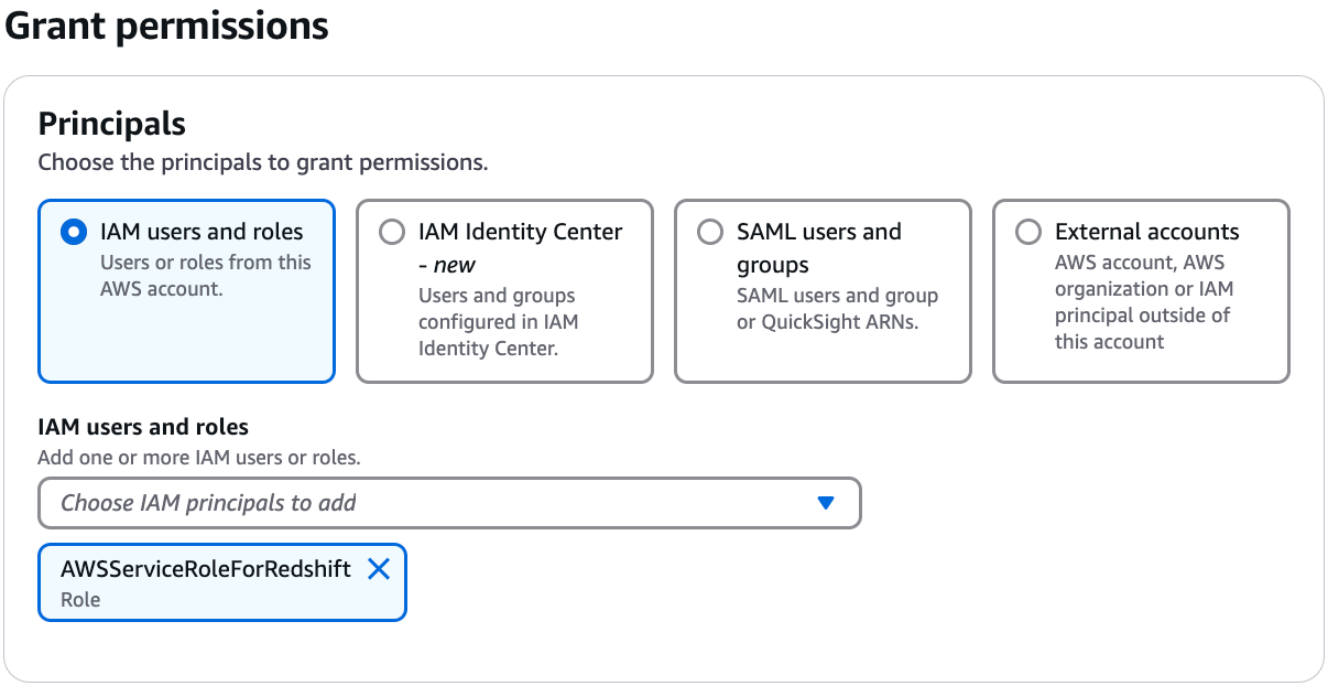
Abra o AWS Lake Formation console em https://console.aws.amazon.com/lakeformation/e faça login como administrador do data lake, criador do banco de dados ou usuário do IAM com permissões concedidas no banco de dados.

Execute um destes procedimentos:

- No painel de navegação, em Permissões, escolha Permissões de dados. Em seguida, escolha Conceder.
- No painel de navegação, selecione Bancos de dados, em Catálogo de dados. Depois, na página Bancos de dados, selecione um banco de dados e, no menu Ações, em Permissões, escolha Conceder.

> **Note**
>
> Você pode conceder permissões em um banco de dados por meio de seu link de recurso. Para fazer isso, na página Bancos de dados, escolha um link de recurso e, no menu Ações, escolha Conceder no destino. Para obter mais informações, consulte Como os links de recursos funcionam no Lake Formation.

2. Na seção Tipo de entidade principal, especifique as entidades principais ou conceda permissões às entidades principais usando atributos.



Usuários e perfis do IAM

Escolha um ou mais usuários ou perfis na lista de usuários e perfis do IAM.

Centro de Identidade do IAM

Selecione um ou mais usuários ou grupos na lista Usuários e grupos. Selecione Adicionar para adicionar mais usuários ou grupos.

Usuários e grupos SAML

Para usuários e grupos do SAML e do Quick, insira um ou mais nomes de recursos da Amazon (ARNs) para usuários ou grupos federados por meio do SAML ou ARNs para usuários ou grupos do Amazon Quick. Pressione Enter após cada ARN.

Para obter informações sobre como construir os ARNs, consulte Concessão e revogação da Lake Formation AWS CLI Comandos da do.

> **Note**
>
> A integração do Lake Formation com o Quick é compatível somente com o Quick Enterprise Edition.

Contas externas

Para Conta da AWS, AWS organização ou diretor do IAM, insira um ou mais IDs de AWS conta, IDs de organização, IDs de unidade organizacional ou ARN válidos para o usuário ou função do IAM. Pressione Enter após cada ID.

O ID da organização consiste em “o-” seguido por 10 a 32 letras minúsculas ou dígitos.

Uma ID de unidade organizacional começa com “ou-” seguida de 4 a 32 letras minúsculas ou dígitos (o ID da raiz que contém a OU). Essa sequência é seguida por um segundo travessão "-" e 8 a 32 letras minúsculas ou dígitos adicionais.

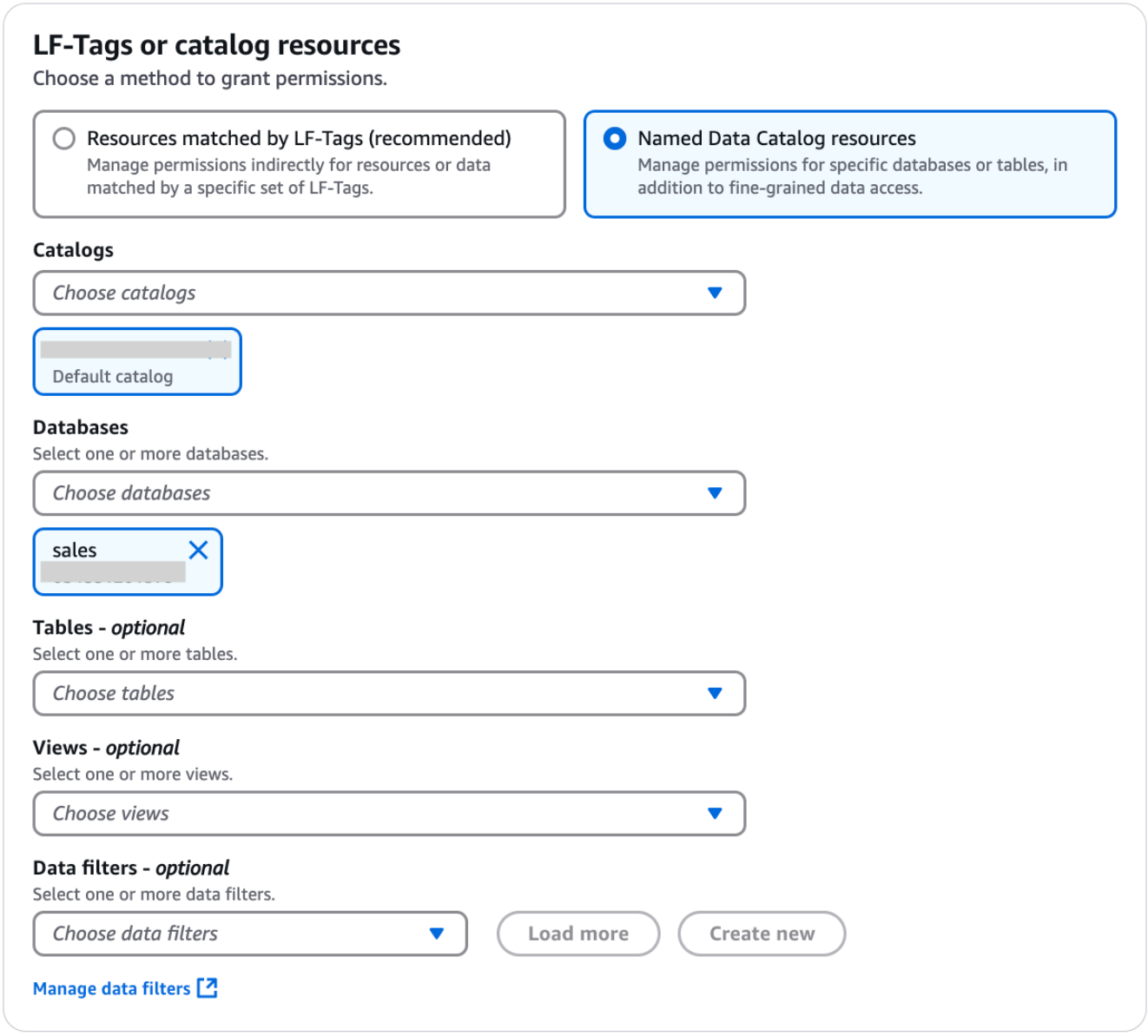
Entidades principais por atributos

Especifique a chave e os valores do atributo. Se você escolher mais de um valor, estará criando uma expressão de atributo com um operador OR. Isso significa que, se algum dos

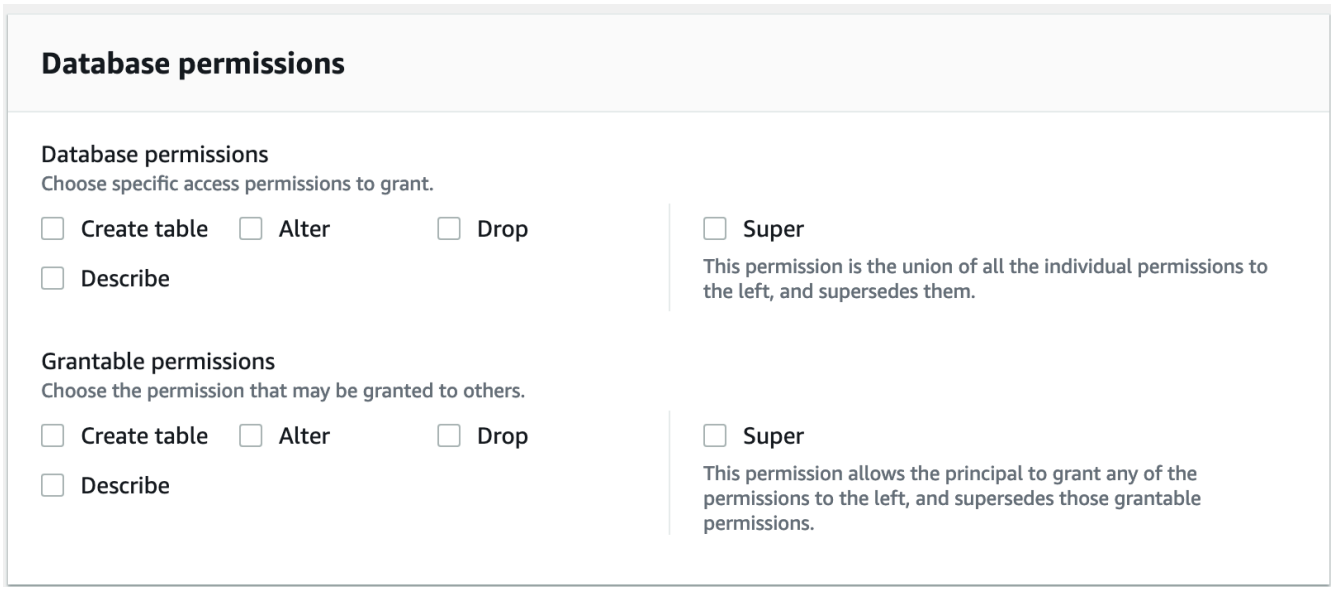
valores da tag de atributo atribuídos a uma função ou usuário do IAM corresponder, eles role/user ganharão permissões de acesso ao recurso.

Escolha o escopo da permissão especificando se você está concedendo permissões a entidades principais com atributos correspondentes na mesma ou em outra conta.

3. Na seção LF-Tags ou recursos do catálogo, escolha Recursos do catálogo de dados nomeados.



4. Escolha um ou mais bancos de dados na lista Banco de dados. Você também pode escolher um ou mais filtros de and/or dados de tabelas.
5. Na seção Permissões, selecione permissões e permissões concedidas. Em Permissões do banco de dados, selecione uma ou mais permissões para conceder.



> **Note**
>
> Depois de conceder `Create Table` ou `Alter` em um banco de dados que tenha uma propriedade de localização que aponta para um local registrado, certifique-se também de conceder permissões de localização de dados no local às entidades principais. Para obter mais informações, consulte Conceder permissões de localização de dados.

6. (Opcional) Em Permissões concedidas, selecione as permissões que o beneficiário da concessão pode conceder a outras entidades principais em sua conta da AWS . Essa opção não é compatível quando você está concedendo permissões a uma entidade principal do IAM a partir de uma conta externa.
7. Selecione Conceder.

AWS CLI

Você pode conceder permissões de banco de dados usando o método de recurso nomeado e o AWS Command Line Interface (AWS CLI).

Para conceder permissões de banco de dados usando o AWS CLI

- Execute um comando `grant-permissions` e especifique um banco de dados ou o catálogo de dados como recurso, dependendo da permissão concedida.

Nos exemplos a seguir, *<account-id>* substitua por um ID de AWS conta válido.

Example– Concessão para criar um banco de dados

Este exemplo concede `CREATE_DATABASE` ao usuário `datalake_user1`. Como o recurso no qual essa permissão é concedida é o catálogo de dados, o comando especifica uma estrutura `CatalogResource` vazia como parâmetro `resource`.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/datalake_user1 --
permissions "CREATE_DATABASE" --resource '{ "Catalog": {}}'
```

Example– Concessão para criar tabelas em um banco de dados designado

O próximo exemplo concede `CREATE_TABLE` no banco de dados `retail` ao usuário `datalake_user1`.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/datalake_user1 --
permissions "CREATE_TABLE" --resource '{ "Database": {"Name":"retail"}}'
```

Example— Concessão a um externo AWS conta com a opção Grant

O próximo exemplo concede `CREATE_TABLE` com a opção de concessão no banco de dados à conta externa `retail` 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333 --permissions "CREATE_TABLE"
 --permissions-with-grant-option "CREATE_TABLE" --resource '{ "Database":
 {"Name":"retail"}}'
```

Example– Concessão a uma organização

O próximo exemplo concede a `ALTER` a opção de concessão no banco de dados `issues` à organização `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- Conceda à AlliamPrincipals na mesma conta

O próximo exemplo concede permissão `CREATE_TABLE` no banco de dados `retail` a todas as entidades principais na mesma conta. Essa opção permite que cada entidade principal da conta crie uma tabela no banco de dados e crie um link de recurso de tabela, permitindo que mecanismos de consulta integrados acessem bancos de dados e tabelas compartilhados. Essa opção é especialmente útil quando uma entidade principal recebe uma concessão entre contas e não tem permissão para criar links de recursos. Nesse cenário, o administrador do data lake pode criar um banco de dados de espaço reservado e conceder permissão `CREATE_TABLE` ao grupo `ALLIAMPrincipal`, permitindo que cada entidade principal do IAM na conta crie links de recursos no banco de dados de espaço reservado.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333:IAMPrincipals
 --permissions "CREATE_TABLE"  --resource '{ "Database":
 {"Name":"temp","CatalogId":"111122223333"}}'
```

Example- Concessão à AlliamPrincipals em uma conta externa

O próximo exemplo concede `CREATE_TABLE` no banco de dados `retail` a todas as entidades principais em uma conta externa. Essa opção permite que cada entidade principal da conta crie uma tabela no banco de dados.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333:IAMPrincipals
 --permissions "CREATE_TABLE"  --resource '{ "Database":
 {"Name":"retail","CatalogId":"123456789012"}}'
```

> **Note**
>
> Depois de conceder `CREATE_TABLE` ou `ALTER` em um banco de dados que tenha uma propriedade de localização que aponta para um local registrado, certifique-se também de conceder permissões de localização de dados no local às entidades principais. Para obter mais informações, consulte Conceder permissões de localização de dados.

Consulte também

- Referência de permissões do Lake Formation
- Conceder permissões em um banco de dados ou tabela compartilhada com sua conta
- Acessar e visualizar tabelas e bancos de dados compartilhados do catálogo de dados

Conceder permissões de tabela usando o método de recurso nomeado

Você pode usar o console do Lake Formation ou AWS CLI conceder permissões do Lake Formation nas tabelas do Catálogo de Dados. Você pode conceder permissões em tabelas individuais ou, com uma única operação de concessão, você pode conceder permissões em todas as tabelas em um banco de dados.

Se você conceder permissões em todas as tabelas em um banco de dados, estará concedendo implicitamente a permissão `DESCRIBE` no banco de dados. Em seguida, o banco de dados aparece na página Bancos de dados no console e é retornado pela operação da API `GetDatabases`. Essa concessão da permissão `DESCRIBE` automática não se aplica quando você usa o controle de acesso por atributo (ABAC). Quando você concede permissões em todas as tabelas em um banco de dados usando atributos, o Lake Formation não concede a permissão `DESCRIBE` implícita ao banco de dados.

Ao escolher `SELECT` como a permissão a ser concedida, você tem a opção de aplicar um filtro de coluna, filtro de linha ou filtro de célula.

Console

As etapas a seguir explicam como conceder permissões de tabela usando o método de recurso nomeado e a página Conceder permissões de data lake no console do Lake Formation. A página está dividida nas seguintes seções:

- Tipos de entidade principal: os usuários, perfis, contas da AWS , organizações ou unidades organizacionais aos quais conceder permissões. Você também pode conceder permissões a entidades principais com atributos correspondentes.
- LF-Tags ou recursos do catálogo — Os bancos de dados, tabelas ou links de recursos nos quais conceder permissões.
- Permissões – As permissões do Lake Formation devem ser concedidas.

> **Note**
>
> Para conceder permissões em um link de recurso de tabela, consulte Como conceder permissões de links de recursos.

1. Abra a página Conceder permissões.

   Abra o AWS Lake Formation console em https://console.aws.amazon.com/lakeformation/e faça login como administrador do data lake, criador da tabela ou usuário que tenha recebido permissões na tabela com a opção de concessão.

   Execute um destes procedimentos:

   - No painel de navegação, escolha Permissões de dados, em Permissões. Em seguida, escolha Conceder.
   - No painel de navegação, selecione Tabelas. Em seguida, na página Tabelas, escolha uma tabela e, no menu Ações, em Permissões, escolha Conceder.

   > **Note**
   >
   > Você pode conceder permissões em uma tabela por meio de seu link de recurso. Para fazer isso, na página Tabelas, escolha um link de recurso e, no menu Ações, escolha Conceder no destino. Para obter mais informações, consulte Como os links de recursos funcionam no Lake Formation.

2. Na seção Tipos de entidade principal, especifique as entidades principais ou entidades principais com atributos correspondentes para conceder permissões.

   Usuários e perfis do IAM

   Escolha um ou mais usuários ou perfis na lista de usuários e perfis do IAM.

   Centro de Identidade do IAM

   Selecione um ou mais usuários ou grupos na lista Usuários e grupos.

Usuários e grupos SAML

Para usuários e grupos SAML e Quick, insira um ou mais Amazon Resource Names (ARNs) para usuários ou grupos federados por meio de SAML ou ARNs para usuários ou grupos Quick. Pressione Enter após cada ARN.

Para obter informações sobre como construir os ARNs, consulte Concessão e revogação da Lake Formation AWS CLI Comandos da do.

> **Note**
>
> A integração do Lake Formation com o Quick é compatível somente com o Quick Enterprise Edition.

Contas externas

Para Conta da AWS , AWS organização ou diretor do IAM, insira um ou mais Conta da AWS IDs válidos, IDs da organização, IDs da unidade organizacional ou o ARN do usuário ou função do IAM. Pressione Enter após cada ID.

O ID da organização consiste em “o-” seguido por 10 a 32 letras minúsculas ou dígitos.

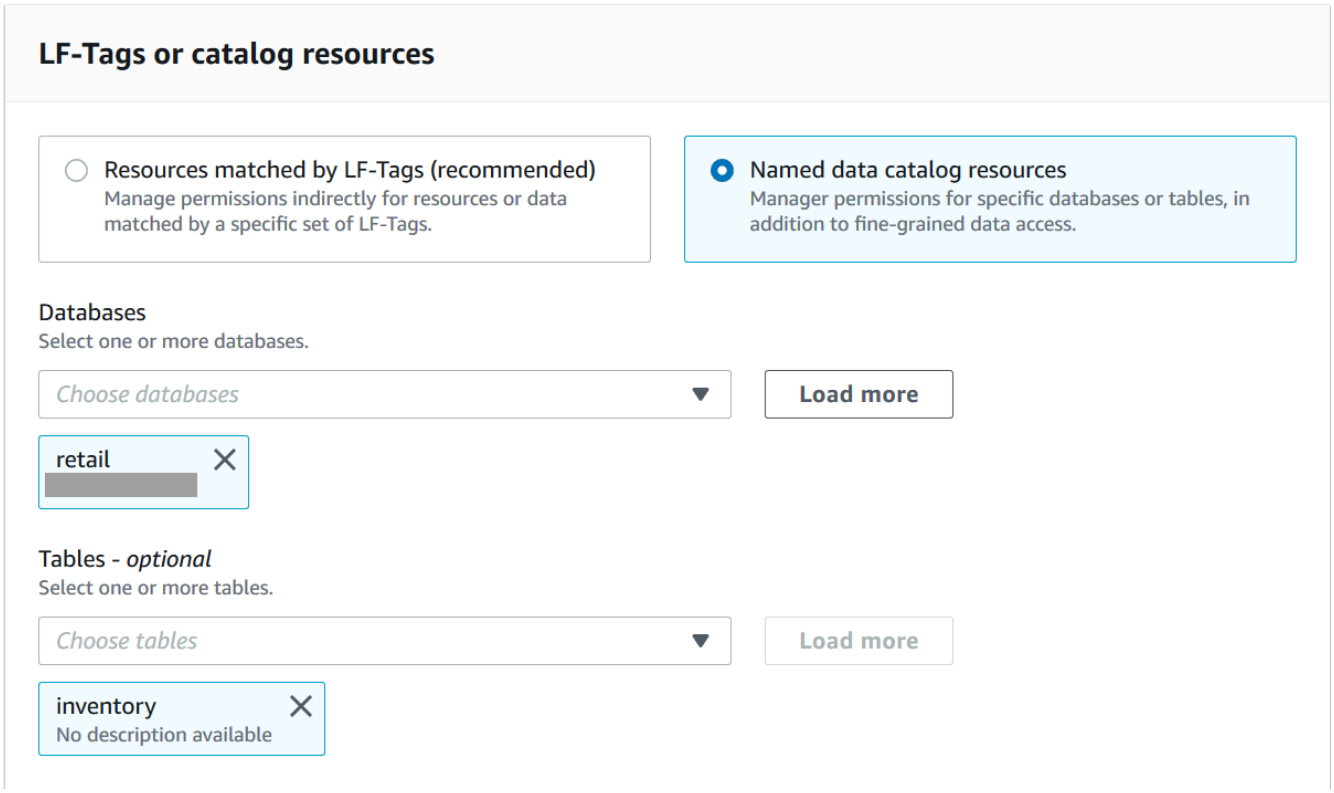
Uma ID de unidade organizacional começa com “ou-” seguida de 4 a 32 letras minúsculas ou dígitos (o ID da raiz que contém a OU). Essa sequência é seguida por um segundo caractere "-" e de 8 a 32 letras minúsculas ou dígitos adicionais.

Entidades principais por atributos

Especifique a chave e os valores do atributo. Se você escolher mais de um valor, estará criando uma expressão de atributo com um operador OR. Isso significa que, se algum dos valores da tag de atributo atribuídos a uma função ou usuário do IAM corresponder, ele role/user ganhará permissões de acesso ao recurso.

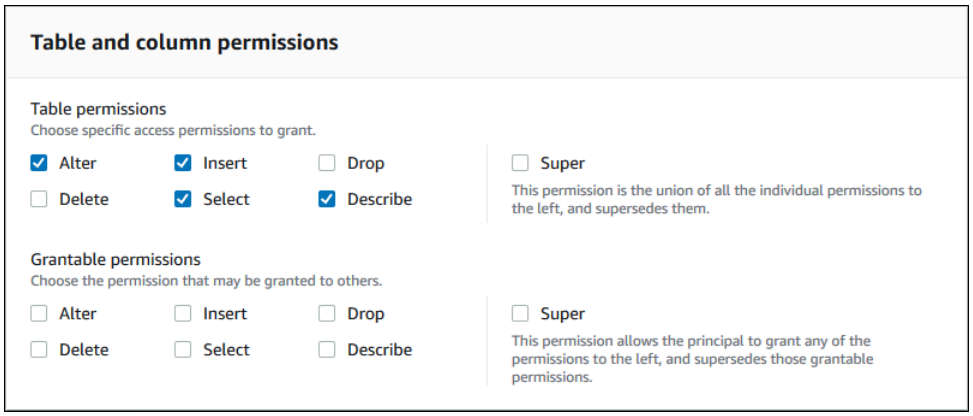
Escolha o escopo da permissão especificando se você está concedendo permissões a entidades principais com atributos correspondentes na mesma ou em outra conta.

3. Na seção LF-Tags ou recursos do catálogo, escolha um banco de dados. Em seguida, escolha uma ou mais tabelas ou Todas as tabelas.

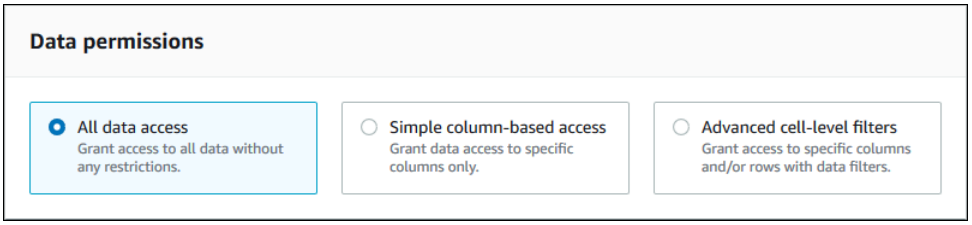


4. Especifique as permissões sem filtragem de dados.

   Na seção Permissões, selecione as permissões da tabela a serem concedidas e, opcionalmente, selecione as permissões que podem ser concedidas.



   Se você conceder Selecionar, a seção Permissões de dados aparecerá abaixo da seção Permissões de tabela e coluna, com a opção Acesso a todos os dados selecionada por padrão. Aceite o padrão.



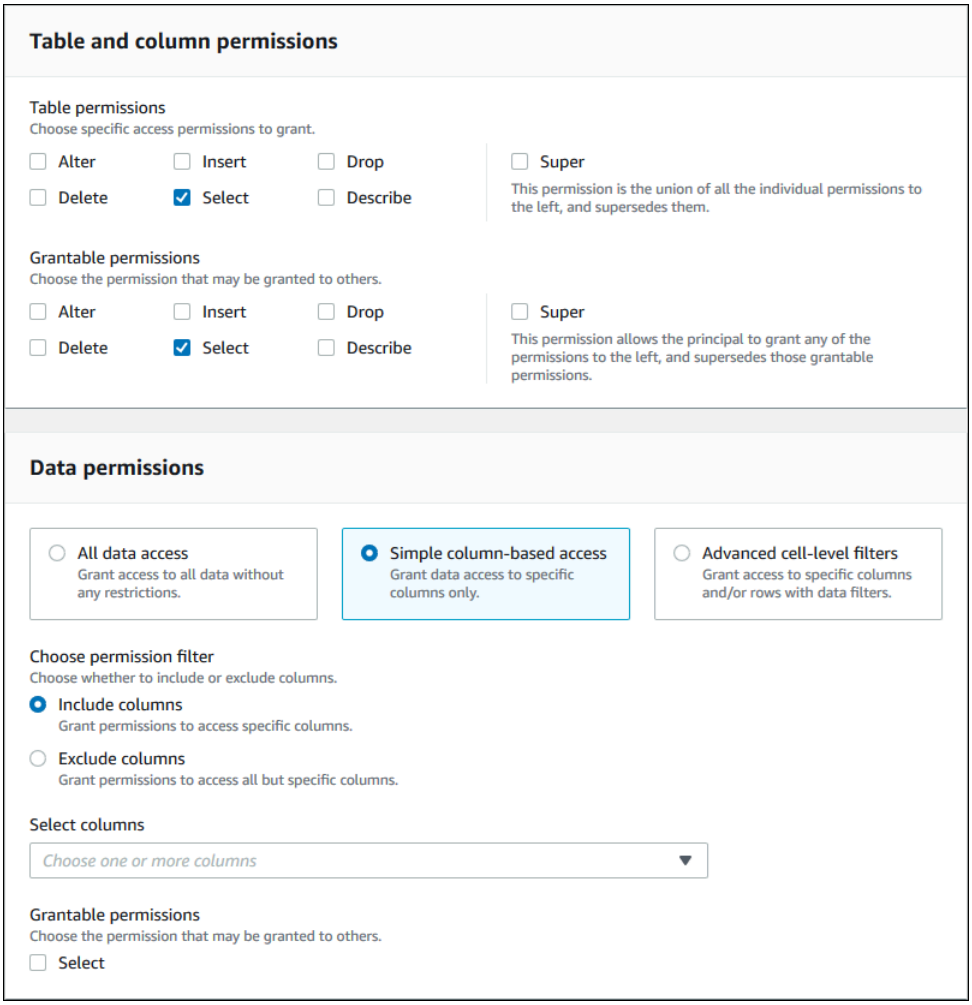
5. Selecione Conceder.
6. Especifique o Select permissão com filtragem de dados

   Selecione a permissão Selecionar. Não selecione nenhuma outra permissão.

   A seção Permissões de dados aparece abaixo da seção Permissões de tabela e coluna.
7. Execute um destes procedimentos:

   - Aplique somente a filtragem simples de colunas.

     1. Escolha Acesso simples baseado em colunas.



2. Escolha se deseja incluir ou excluir colunas e, em seguida, escolha as colunas a serem incluídas ou excluídas.

   Somente listas de inclusão são suportadas ao conceder permissões a uma AWS conta ou organização externa.

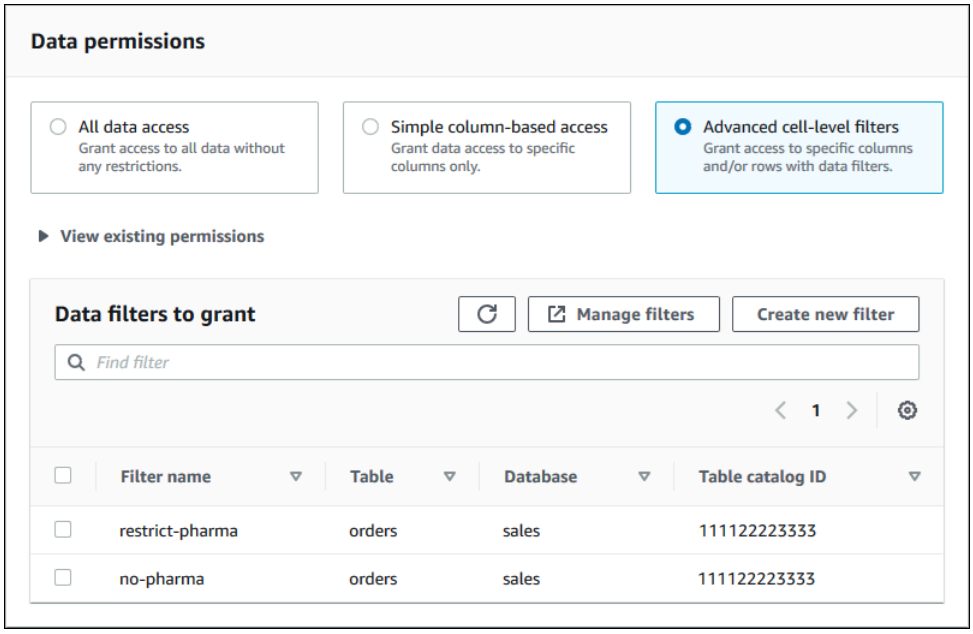
3. (Opcional) Em Permissões concedidas, ative a opção de concessão para a permissão Selecionar.

   Se você incluir a opção de concessão, o destinatário da concessão poderá conceder permissões somente nas colunas que você conceder a ele.

> **Note**
>
> Você também pode aplicar a filtragem de colunas somente criando um filtro de dados que especifique um filtro de coluna e especifique todas as linhas como filtro de linha. No entanto, isso requer mais etapas.

- Aplique filtragem de coluna, linha ou célula.

  1. Escolha Filtros avançados em nível de célula.



  2. (Opcional) Expanda Visualizar permissões existentes.
  3. (Opcional) Escolha Criar novo filtro.
  4. (Opcional) Para ver detalhes dos filtros listados, criar novos filtros ou excluir filtros existentes, escolha Gerenciar filtros.

A página Filtros de dados é aberta em uma nova janela do navegador.

Quando terminar de acessar a página Filtros de dados, retorne à página Conceder permissões e, se necessário, atualize a página para ver os novos filtros de dados que você criou.

5. Selecione um ou mais filtros de dados a serem aplicados à concessão.

> **Note**
>
> Se não houver filtros de dados na lista, isso significa que nenhum filtro de dados foi criado para a tabela selecionada.

8. Selecione Conceder.

AWS CLI

Você pode conceder permissões de tabela usando o método de recurso nomeado e o AWS Command Line Interface (AWS CLI).

Para conceder permissões de tabela usando o AWS CLI

- Execute um comando `grant-permissions` e especifique uma tabela como recurso.

Example– Concessão em uma única tabela - sem filtragem

O exemplo a seguir concede `SELECT` e `ALTER` ao usuário `datalake_user1` na AWS conta 1111-2222-3333 na tabela no banco de dados. `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**
>
> Se você conceder a permissão `ALTER` em uma tabela que tem seus dados subjacentes em um local registrado, certifique-se de também conceder permissões de localização de

dados no local às entidades principais. Para obter mais informações, consulte Conceder permissões de localização de dados.

Example– Conceder em todas as tabelas com a opção Conceder - sem filtragem

O próximo exemplo concede SELECT com a opção de concessão em todas as tabelas no banco de dados `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– Grant com filtragem simples de colunas

O próximo exemplo concede a SELECT um subconjunto de colunas na tabela `persons`. Ele usa filtragem de coluna simples.

```
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– Conceda com um filtro de dados

Este exemplo concede SELECT na tabela `orders` e aplica o filtro de dados `restrict-pharma`.

```
aws lakeformation grant-permissions --cli-input-json file://grant-params.json
```

A seguir está o conteúdo do arquivo `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"]
}
```

Consulte também

- Visão geral das permissões do Lake Formation
- Filtragem de dados e segurança por célula no Lake Formation
- Referência de personas e permissões do IAM do Lake Formation
- Como conceder permissões de links de recursos
- Acessar e visualizar tabelas e bancos de dados compartilhados do catálogo de dados

Conceder permissões em visualizações usando o método de recurso nomeado

As etapas a seguir explicam como conceder permissões em visualizações usando o método de recurso nomeado e a página Conceder permissões. A página está dividida nas seguintes seções:

- Tipos principais: usuários, funções, usuários e grupos do IAM Identity Center Contas da AWS, organizações ou unidades organizacionais para conceder permissões. Você também pode conceder permissões a entidades principais com atributos correspondentes.
- LF-Tags ou recursos do catálogo — Os bancos de dados, tabelas, visualizações ou links de recursos nos quais conceder permissões.
- Permissões: as permissões de data lake a serem concedidas.

Abra as Conceder permissões page

1. Abra o AWS Lake Formation console em https://console.aws.amazon.com/lakeformation/e faça login como administrador do data lake, criador do banco de dados ou usuário do IAM com permissões concedidas no banco de dados.
2. Execute um destes procedimentos:

- No painel de navegação, em Permissões, escolha Permissões de dados. Em seguida, escolha Conceder.
- No painel de navegação, selecione Visualizações em Catálogo de dados. Depois, na página Visualizações, selecione uma visualização e, no menu Ações, em Permissões, escolha Conceder.

> **Note**
>
> É possível conceder permissões em uma visualização por meio do link de recurso. Para fazer isso, na página Visualizações, escolha um link de recurso e, no menu Ações, selecione Conceder no destino. Para obter mais informações, consulte Como os links de recursos funcionam no Lake Formation.

Especificar os tipos de entidade principal

Na seção Tipo de entidade principal, escolha Entidades principais ou Entidades principais por atributos. Se você escolher Entidades principais, as seguintes opções estarão disponíveis:

Usuários e perfis do IAM

Escolha um ou mais usuários ou perfis na lista de usuários e perfis do IAM.

Centro de Identidade do IAM

Selecione um ou mais usuários ou grupos na lista Usuários e grupos.

Usuários e grupos SAML

Para usuários e grupos do SAML e do Quick, insira um ou mais nomes de recursos da Amazon (ARNs) para usuários ou grupos federados por meio do SAML ou ARNs para usuários ou grupos do Amazon Quick. Pressione Enter após cada ARN.

Para obter informações sobre como construir os ARNs, consulte Concessão e revogação da Lake Formation AWS CLI Comandos da do.

> **Note**
>
> A integração do Lake Formation com o Quick é compatível somente com o Quick Enterprise Edition.

Contas externas

Para Conta da AWS, AWS organização ou diretor do IAM, insira um ou mais IDs de AWS conta, IDs de organização, IDs de unidade organizacional ou ARN válidos para o usuário ou função do IAM. Pressione Enter após cada ID.

O ID da organização consiste em “o-” seguido por 10 a 32 letras minúsculas ou dígitos.

Uma ID de unidade organizacional começa com “ou-” seguida de 4 a 32 letras minúsculas ou dígitos (o ID da raiz que contém a OU). Essa sequência é seguida por um segundo travessão "-" e 8 a 32 letras minúsculas ou dígitos adicionais.

> **Consulte também**
>
> - Acessar e visualizar tabelas e bancos de dados compartilhados do catálogo de dados

Entidades principais por atributos

Especifique a chave e os valores do atributo. Se você escolher mais de um valor, estará criando uma expressão de atributo com um operador OR. Isso significa que, se algum dos valores da tag de atributo atribuídos a uma função ou usuário do IAM corresponder, ele role/user ganhará permissões de acesso ao recurso.

Escolha o escopo da permissão especificando se você está concedendo permissões a entidades principais com atributos correspondentes na mesma ou em outra conta.

Especificar as visualizações

Na seção LF-Tags ou recursos do catálogo, escolha uma ou mais visualizações para conceder permissões.

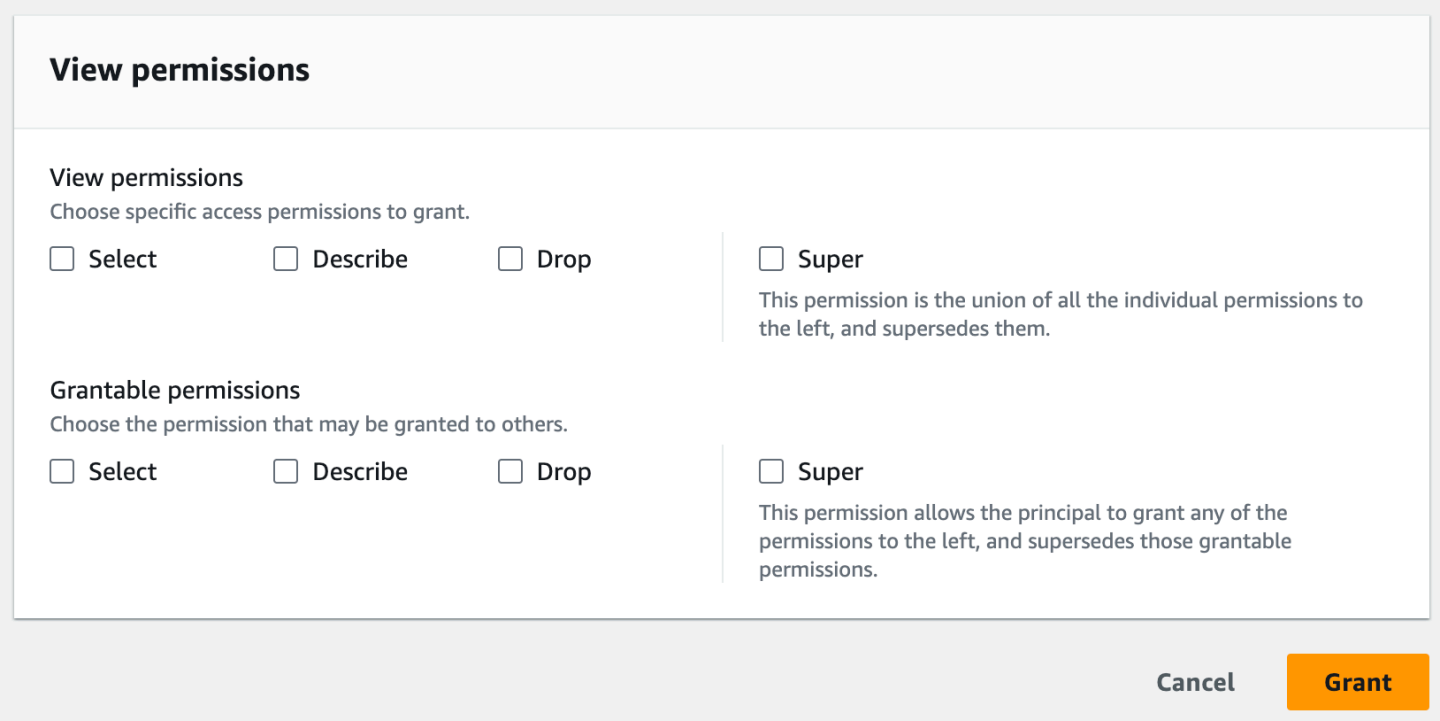
1. Escolha Recursos do catálogo de dados nomeado.

2. Selecione uma ou mais visualizações na lista Visualizações. Você também pode escolher um ou mais catálogos, bancos de dados, tabelas e filtros de and/or dados.

   Conceder permissões de data lake a `All tables` em um banco de dados resultará em permissões do favorecido em todas as tabelas e visualizações do banco de dados.

Especifique as permissões

Na seção Permissões, selecione permissões e permissões concedidas.



1. Em Visualizar permissões, selecione uma ou mais permissões a serem concedidas.
2. (Opcional) Em Permissões que podem ser concedidas, selecione as permissões que o beneficiário pode conceder a outras entidades principais na Conta da AWS. Essa opção não é compatível quando você está concedendo permissões a uma entidade principal do IAM a partir de uma conta externa.
3. Selecione Conceder.

Consulte também

- Referência de permissões do Lake Formation

- Conceder permissões em um banco de dados ou tabela compartilhada com sua conta

Controle de acesso baseado em tags do Lake Formation

O controle de acesso baseado em tags do Lake Formation (LF-TBAC) é uma estratégia de autorização que define permissões com base em atributos. Em Lake Formation, esses atributos são chamados LF-Tags. Você pode LF-Tags se conectar aos recursos do Catálogo de Dados e conceder permissões aos diretores do Lake Formation sobre esses recursos usando esses LF-Tags recursos. O Lake Formation permite operações nesses recursos quando a entidade principal possui acesso concedido a um valor de tag correspondente ao valor da tag do recurso.

LF-TBAC é útil em ambientes que estão crescendo rapidamente e ajuda em situações em que o gerenciamento de políticas se torna complicado.

LF-TBAC é o método recomendado para conceder permissões ao Lake Formation quando há um grande número de objetos do Catálogo de Dados, incluindo catálogos federados, bancos de dados, tabelas e visualizações. O Lake Formation oferece suporte ao controle de acesso baseado em tags para catálogos federados de tabelas, Amazon Redshift data warehouses e fontes de dados federadas do Amazon S3, como Amazon DynamoDB SQL Server e Snowflake.

> **Note**
>
> As tags IAM não são iguais LF-Tags a. Essas etiquetas não são intercambiáveis. LF-Tagssão usadas para conceder permissões do Lake Formation e as tags do IAM são usadas para definir políticas do IAM.

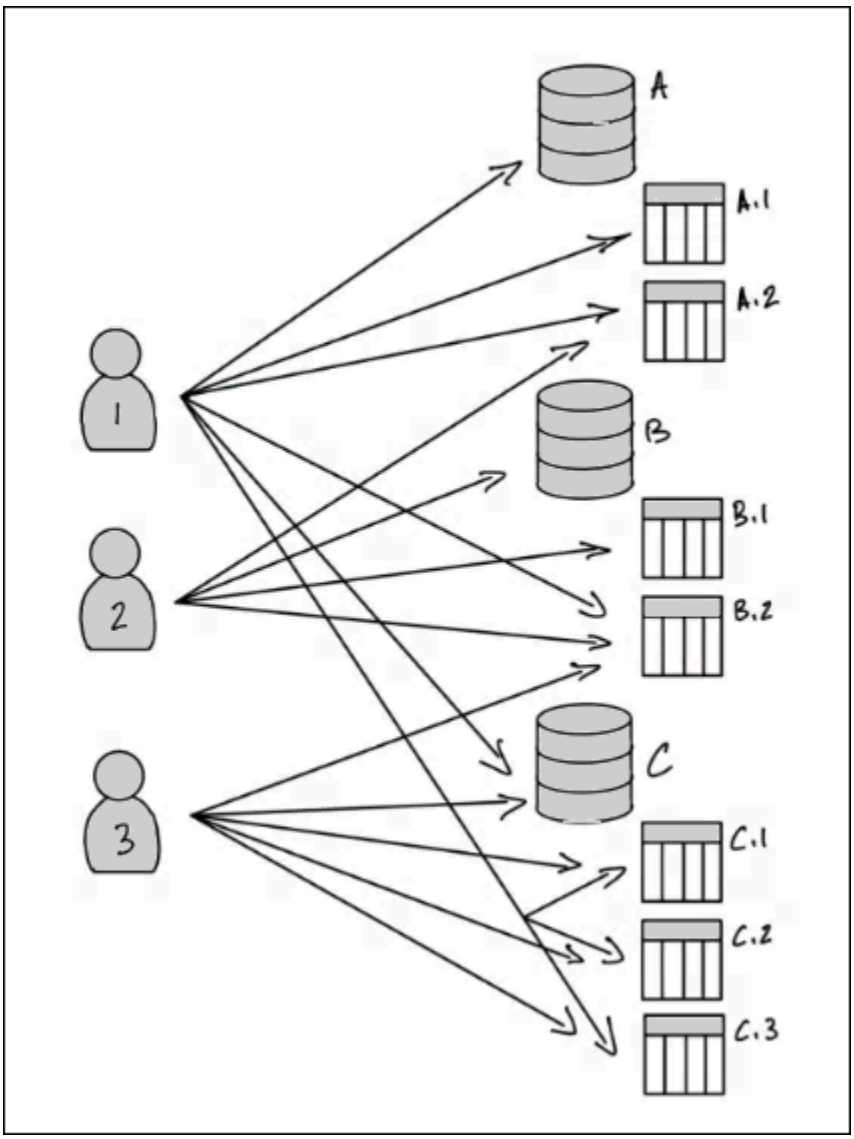
Como funciona o controle de acesso baseado em tags do Lake Formation

Cada um LF-Tag é um par de valores-chave, como `department=sales` ou. `classification=restricted` Uma chave pode ter vários valores definidos, como `department=sales,marketing,engineering,finance`.

Para usar o LF-TBAC método, os administradores e engenheiros de dados do data lake realizam as seguintes tarefas.

Tarefa	Detalhes da tarefa
1. Defina as propriedades e os relaciona mentos de LF-Tags.	-
2. Crie os LF-Tag criadores em Lake Formation.	Adicionando LF-Tag criadores
3. Crie o LF-Tag em Lake Formation.	Criando LF-Tags
4. Atribua LF-Tags aos recursos do Catálogo de Dados.	Atribuição LF-Tags aos recursos do Catálogo de Dados
5. Conceda permissões a outros diretores para LF-Tags atribuir recursos, opcionalm ente com a opção de concessão.	Gerenciando permissões de LF-Tag valor
6. Conceda LF-Tag expressões aos diretores, opcionalmente com a opção de concessão.	Concedendo permissões de data lake usando o método LF-TBAC
7. (Recomendado) Depois de verificar se os diretores têm acesso aos recursos corretos por meio do LF-TBAC método, revogue as permissões concedidas usando o método de recurso nomeado.	-

Considere o caso em que você deve conceder permissões a três entidade principais em três bancos de dados e sete tabelas.



Para obter as permissões indicadas no diagrama anterior usando o método de recurso nomeado, você precisaria fazer 17 concessões, da seguinte forma (em pseudocódigo).

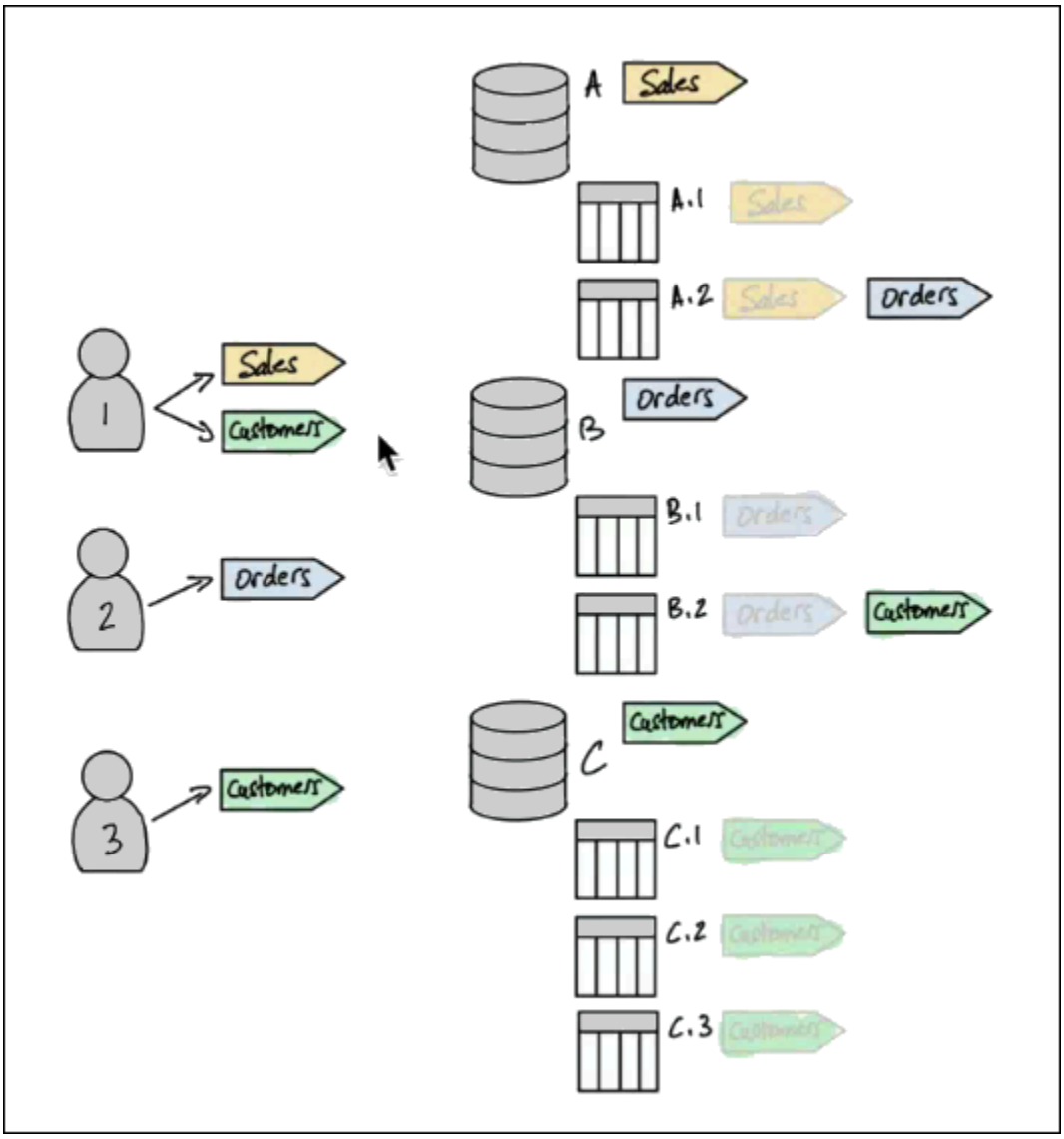
```
GRANT CREATE_TABLE ON Database A TO PRINCIPAL 1
GRANT SELECT, INSERT ON Table A.1 TO PRINCIPAL 1
GRANT SELECT, INSERT ON Table A.2 TO PRINCIPAL 1
GRANT SELECT, INSERT ON Table B.2 TO PRINCIPAL 1
...
GRANT SELECT, INSERT ON Table A.2 TO PRINCIPAL 2
GRANT CREATE_TABLE ON Database B TO PRINCIPAL 2
...
GRANT SELECT, INSERT ON Table C.3 TO PRINCIPAL 3
```

Agora, considere como você concederia permissões usando LF-TBAC o. O diagrama a seguir indica que você atribuiu LF-Tags a bancos de dados e tabelas e concedeu permissões LF-Tags aos principais.

Neste exemplo, LF-Tags representam áreas do data lake que contêm análises para diferentes módulos de um pacote de aplicativos de planejamento de recursos corporativos (ERP). Você pode usá-las para controlar o acesso aos dados de analytics dos vários módulos. Todos LF-Tags têm a chave `module` e os valores possíveis `SalesOrders`, `Customers` e. Um exemplo LF-Tag se parece com isso:

```
module=Sales
```

O diagrama mostra somente os LF-Tag valores.



Atribuições de tags aos recursos e herança do catálogo de dados

As tabelas herdam LF-Tags dos bancos de dados e as colunas herdam LF-Tags das tabelas. Os valores herdados podem ser substituídos. No diagrama anterior, os esmaecidos são LF-Tags herdados.

Por causa da herança, o administrador do data lake precisa fazer somente as cinco LF-Tag atribuições a seguir aos recursos (em pseudocódigo).

```
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
```

Concessões de tags a entidade principais

Depois de atribuir LF-Tags aos bancos de dados e tabelas, o administrador do data lake deve fazer apenas quatro concessões LF-Tags aos diretores, da seguinte forma (em pseudocódigo).

```
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
```

Agora, um diretor com o `module=Sales` LF-Tag pode acessar os recursos do Catálogo de Dados com o `module=Sales` LF-Tag (por exemplo, banco de dados A), um diretor com o `module=Customers` LF-Tag pode acessar recursos com o `module=Customers` LF-Tag e assim por diante.

Os comandos de concessão anteriores estão incompletos. Isso ocorre porque, embora indiquem, por meio LF-Tags dos recursos do Catálogo de Dados, que os diretores têm permissões, eles não indicam exatamente quais permissões do Lake Formation (por exemplo`SELECT`,`ALTER`) os diretores têm sobre esses recursos. Portanto, os comandos de pseudocódigo a seguir são uma representação mais precisa de como as permissões do Lake Formation são concedidas nos recursos do Catálogo de Dados por meio de. LF-Tags

```
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
```

Reunir tudo: permissões resultantes em recursos

Considerando a LF-Tags atribuição aos bancos de dados e tabelas no diagrama anterior e a LF-Tags concedida aos diretores no diagrama, a tabela a seguir lista as permissões do Lake Formation que os diretores têm nos bancos de dados e tabelas.

Entidade principal	Permissões concedidas por meio de LF-Tags
Entidade principal 1	• `CREATE_TABLE` no banco de dados A • `SELECT`, `INSERT` na mesa A.1 • `SELECT`, `INSERT` na mesa B.2 • `CREATE_TABLE` no banco de dados C • `SELECT`, `INSERT` na mesa C.1 • `SELECT`, `INSERT` na mesa C.2 • `SELECT`, `INSERT` na mesa C.3
Entidade principal 2	• `SELECT`, `INSERT` na mesa A.2 • `CREATE_TABLE` no banco de dados B • `SELECT`, `INSERT` na mesa B.1
Entidade principal 3	• `SELECT`, `INSERT` na mesa B.2 • `CREATE_TABLE` no banco de dados C • `SELECT`, `INSERT` na mesa C.1 • `SELECT`, `INSERT` na mesa C.2 • `SELECT`, `INSERT` na mesa C.3

Conclusão

Neste exemplo simples, usando cinco operações de atribuição e oito operações de concessão, o administrador do data lake conseguiu especificar 17 permissões. Quando há dezenas de bancos de

dados e centenas de tabelas, a vantagem do LF-TBAC método sobre o método de recurso nomeado fica clara. No caso hipotético da necessidade de conceder a todos os principais acesso a todos os recursos, e onde `n(P)` é o número de entidades principais e `n(R)` o número de recursos:

- Com o método de recurso nomeado, o número de concessões necessárias é `n(P)` × `n(R)`.
- Com o LF-TBAC método, usando um único LF-Tag, o total do número de doações para diretores e atribuições para recursos é `n(P)` +. `n(R)`

> Consulte também
>
> - Gerenciamento do controle LF-Tags de acesso aos metadados
> - Concedendo permissões de data lake usando o método LF-TBAC

Tópicos

- Gerenciamento do controle LF-Tags de acesso aos metadados
- Gerenciando LF-Tag expressões para controle de acesso a metadados
- Gerenciando permissões de LF-Tag valor

Gerenciamento do controle LF-Tags de acesso aos metadados

Para usar o método de controle de acesso (LF-TBAC) baseado em tags do Lake Formation para proteger objetos do Catálogo de Dados, como catálogos, bancos de dados, tabelas, visualizações e colunas, você os cria LF-Tags, atribui aos recursos e concede LF-Tag permissões aos diretores.

Antes de LF-Tags atribuir objetos do Catálogo de Dados ou conceder permissões aos principais, você precisa definir LF-Tags. Somente um administrador de data lake ou um diretor com permissões de LF-Tag criador pode criar LF-Tags.

LF-Tag criadores

LF-Tag O criador é um diretor não administrador que tem permissões para criar e gerenciar LF-Tags. Os administradores do Data Lake podem adicionar LF-Tag criadores usando o console ou a CLI do Lake Formation. LF-Tag os criadores têm permissões implícitas do Lake Formation para atualizar e excluir LF-Tags, LF-Tags atribuir recursos e conceder LF-Tag permissões e permissões de LF-Tag valor a outros diretores.

Com as funções de LF-Tag criador, os administradores do data lake podem delegar tarefas de gerenciamento de tags, como criar e atualizar chaves e valores de tags, a diretores que não são administradores. Os administradores do Data Lake também podem conceder permissões LF-Tag `Create LF-Tag` concedidas aos criadores. Em seguida, o LF-Tag criador pode conceder a permissão para criar LF-Tags a outros diretores.

Você pode conceder dois tipos de permissões em LF-Tags:

- LF-Tag permissões - `Create LF-TagAlter`, `Drop` e. Essas permissões são necessárias para criar, atualizar e excluir LF-Tags.

  Os administradores e LF-Tag criadores de data lake têm implicitamente essas permissões no data LF-Tags que criam e podem conceder essas permissões explicitamente aos diretores para gerenciar tags no data lake.
- LF-Tag permissões de pares de valores-chave - `AssignDescribe`, e. `Grant with LF-Tag expressions` Essas permissões são necessárias para atribuir LF-Tags aos objetos do Catálogo de Dados e para conceder permissões sobre os recursos aos diretores usando o controle de acesso baseado em tags do Lake Formation. LF-Tag os criadores recebem implicitamente essas permissões ao criar LF-Tags.

Depois de receber a `Create LF-Tag` permissão e criar com sucesso LF-Tags, o LF-Tag criador pode LF-Tags atribuir recursos e conceder LF-Tag permissões (`Create LF-Tag`,`Alter`,`Drop`, e) a outros diretores não administrativos para gerenciar tags no data lake. Você pode gerenciar LF-Tags usando o console do Lake Formation, a API ou o AWS Command Line Interface (AWS CLI).

> **Note**
>
> Os administradores do Data Lake têm permissões implícitas do Lake Formation para criar, atualizar e excluir LF-Tags, LF-Tags atribuir recursos e conceder LF-Tag permissões aos diretores.

Para conhecer as práticas recomendadas e as considerações, consulte Considerações e práticas recomendadas de controle de acesso com base em tags do Lake Formation.

Tópicos

- Adicionando LF-Tag criadores
- Criando LF-Tags

- Atualizando LF-Tags
- Excluindo LF-Tags
- Listagem LF-Tags
- Atribuição LF-Tags aos recursos do Catálogo de Dados
- Visualização LF-Tags atribuída a um recurso
- Visualizando os recursos aos quais um LF-Tag está atribuído
- Ciclo de vida de um LF-Tag
- Comparação do controle de acesso baseado em tags do Lake Formation com o controle de acesso baseado em atributos do IAM

Consulte também

- Gerenciando permissões de LF-Tag valor
- Concedendo permissões de data lake usando o método LF-TBAC
- Controle de acesso baseado em tags do Lake Formation

Adicionando LF-Tag criadores

Por padrão, os administradores do data lake podem criar, atualizar e excluir LF-Tags, atribuir tags aos objetos do catálogo de dados e conceder permissões de tag aos principais. Se você quiser delegar as operações de criação e gerenciamento de tags a diretores que não são administradores, o administrador do data lake pode criar funções de LF-Tag criador e `Create LF-Tag` conceder permissão ao Lake Formation para as funções. Com a `Create LF-Tag` permissão concedida, LF-Tag os criadores podem delegar tarefas de criação e manutenção de tags a outros diretores não administrativos.

Para que os administradores do data lake atribuam LF-Tags aos recursos do Catálogo de Dados, eles precisam conceder a si mesmos permissões de associação LF-Tags que não foram criadas por eles.

Note

Cross-account as concessões de permissão podem incluir somente `Describe` e `Associate` permissões. Você não pode conceder permissões `Create LF-Tag`, `Drop`,

`Alter` e `Grant with LFTag expressions` a entidades principais em uma conta diferente.

Tópicos

- Permissões do IAM necessárias para criar LF-Tags
- Adicionar LF-Tag criadores

Consulte também

- Gerenciando permissões de LF-Tag valor
- Concedendo permissões de data lake usando o método LF-TBAC
- Controle de acesso baseado em tags do Lake Formation

Permissões do IAM necessárias para criar LF-Tags

Você deve configurar as permissões para permitir que um diretor do Lake Formation crie LF-Tags. Adicione a seguinte declaração à política de permissões do diretor que precisa ser um LF-Tag criador.

Note

Embora os administradores do data lake tenham permissões implícitas do Lake Formation para criar, atualizar e excluir LF-Tags, LF-Tags atribuir recursos e conceder LF-Tags aos diretores, os administradores do data lake também precisam das seguintes permissões do IAM.

Para obter mais informações, consulte Referência de personas e permissões do IAM do Lake Formation.

```
{
"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"
    ]
}
```

Os diretores que LF-Tags atribuem recursos e concedem LF-Tags aos diretores devem ter as mesmas permissões, exceto as permissões `CreateLFTag`Update`LFTag`, e. `DeleteLFTag`

Adicionar LF-Tag criadores

Um LF-Tag criador pode criar uma LF-Tag, atualizar a chave e os valores da tag, excluir tags, associar tags aos recursos do Catálogo de Dados e conceder permissões sobre os recursos do Catálogo de Dados aos diretores usando o LF-TBAC método. O LF-Tag criador também pode conceder essas permissões aos diretores.

Você pode criar papéis de LF-Tag criador usando o AWS Lake Formation console, a API ou o AWS Command Line Interface (AWS CLI).

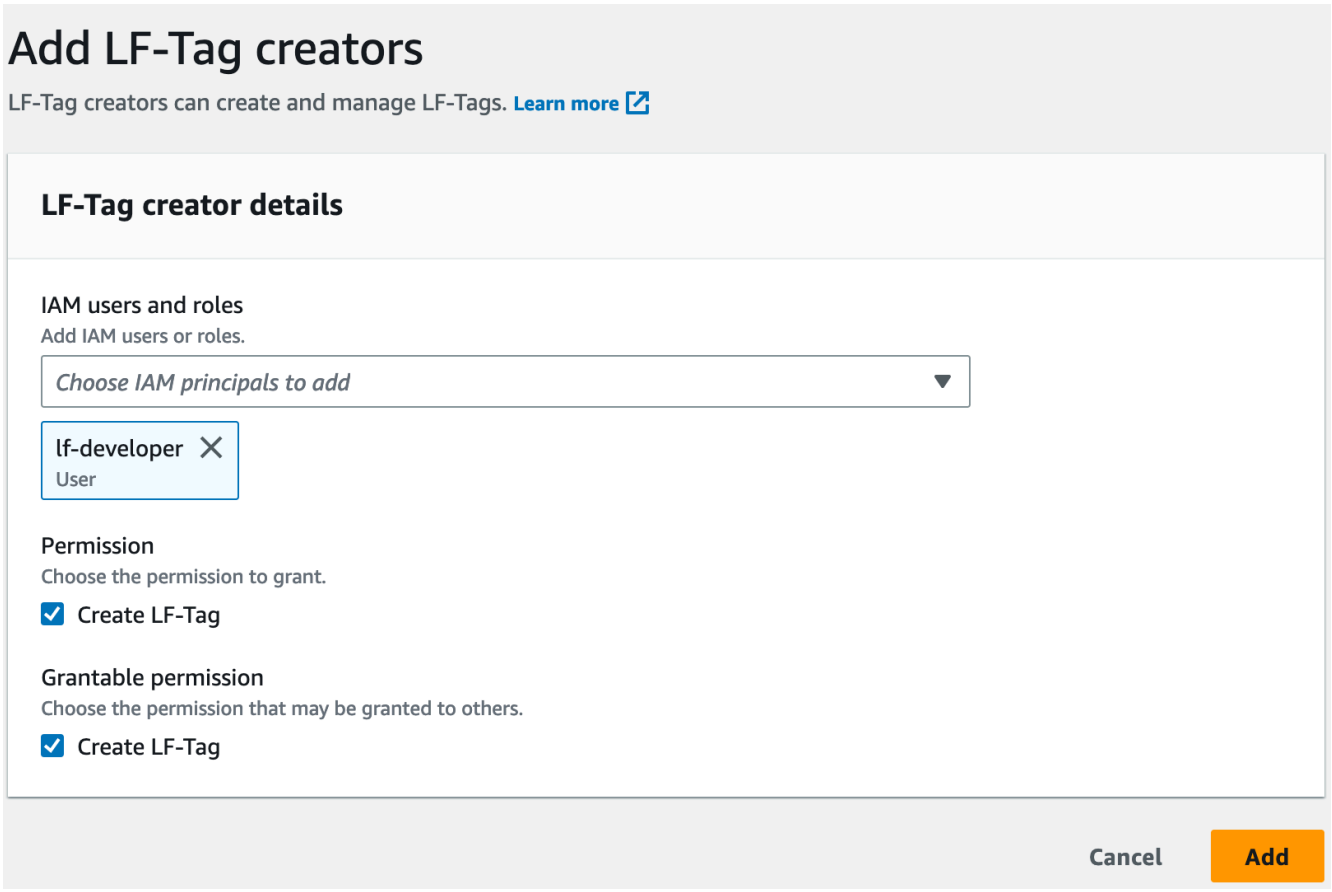
console

Para adicionar um LF-Tag criador

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

   Faça login como administrador de data lake.
2. No painel de navegação, em Permissões, escolha LF-Tags e permissões.

   Na página LF-Tags e permissões, escolha a seção LF-Tag Criadores e escolha Adicionar LF-Tag criadores.



3. Na página Adicionar LF-Tag criadores, escolha uma função do IAM ou um usuário que tenha as permissões necessárias para criar LF-Tags.
4. Ativar caixa de seleção de permissão `Create LF-Tag`.
5. (Opcional) Para permitir que as entidades principais selecionadas concedam a permissão `Create LF-Tag` às entidades principais, escolha a permissão `Create LF-Tag` concedível.
6. Escolha Adicionar.

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"
    ]
}
```

A seguir estão as permissões disponíveis para uma função de LF-Tag criador:

Permissão	Description
`Drop`	Um diretor com essa permissão em um LF-Tag pode excluir um LF-Tag do data lake. O diretor obtém `Describe` permissão implícita em todos os valores de tag de um LF-Tag recurso.
`Alter`	Um diretor com essa permissão em um LF-Tag pode adicionar ou remover o valor da tag de um LF-Tag. O diretor obtém `Alter` permissão implícita em todos os valores de tag de a. LF-Tag
`Describe`	Um diretor com essa permissão em um LF-Tag pode visualizar o LF-Tag e seus valores ao LF-Tags atribuir recursos ou conceder permissões em LF-Tags. Você pode conceder `Describe` em todos os valores-chave ou em valores específicos.
`Associate`	Um diretor com essa permissão em um LF-Tag pode atribuí-lo LF-Tag a um recurso do Catálogo de Dados. Conceder `Associate` concede implicitamente `Describe`.
`Grant with LF-Tag expression`	Um diretor com essa permissão em um LF-Tag pode conceder permissões sobre os recursos de um catálogo de dados usando a LF-Tag chave e os valores. Conceder `Grant with LF-Tag expression` concede implicitamente `Describe`.

Essas permissões são concedidas. Uma entidade principal que tenha recebido essas permissões com a opção de concessão pode concedê-las a outras entidades principais.

Criando LF-Tags

Todos LF-Tags devem ser definidos em Lake Formation antes de serem usados. A LF-Tag consiste em uma chave e em um ou mais valores possíveis para a chave.

Depois que o administrador do data lake configurar as permissões necessárias do IAM e as permissões do Lake Formation para a função de LF-Tag criador, o diretor pode criar uma LF-Tag. O LF-Tag criador obtém permissão implícita para atualizar ou remover qualquer valor de tag do LF-Tag e excluir o. LF-Tag

Você pode criar LF-Tags usando o AWS Lake Formation console, a API ou o AWS Command Line Interface (AWS CLI).

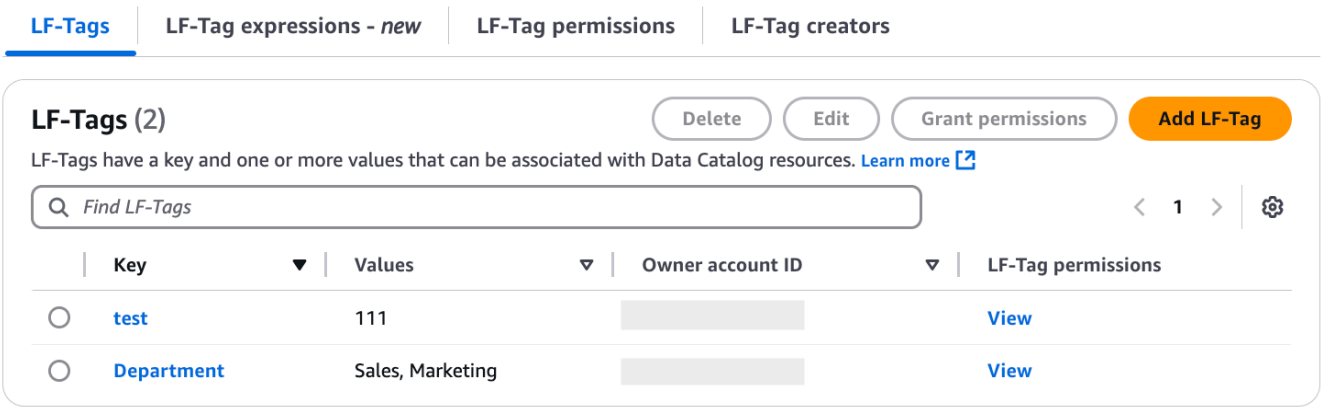
Console

Para criar um LF-Tag

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

   Faça login como diretor com permissões de LF-Tag criador ou como administrador do data lake.
2. No painel de navegação, em Permissões LF-Tags e permissões, escolha LF-Tags.

   A página LF-Tags é exibida.



3. Escolha Adicionar LF-Tag.
4. Na caixa LF-Tag de diálogo Adicionar, insira uma chave e um ou mais valores.

   Cada chave deve ter pelo menos um valor. Para inserir vários valores, insira uma lista delimitada por vírgulas e pressione Enter ou insira um valor por vez e escolha Adicionar após cada um. O número máximo de aliases por usuário é de 1000.

5. Escolha Adicionar Tag.

AWS CLI

Para criar um LF-Tag

- Insira um comando `create-lf-tag`.

  O exemplo a seguir cria um LF-Tag com chave `module` `Customers` e valores `Orders` e.

```
aws lakeformation create-lf-tag --tag-key module --tag-values Customers Orders
```

Como criador da tag, o diretor obtém `Alter` permissão para isso LF-Tag e pode atualizar ou remover qualquer valor da tag LF-Tag. O diretor LF-Tag criador também pode conceder `Alter` permissão a outro diretor para atualizar e remover os valores das tags LF-Tag.

Atualizando LF-Tags

Você LF-Tag atualiza uma na qual tem `Alter` permissão adicionando ou excluindo valores de chave permitidos. Você não pode mudar a LF-Tag chave. Para alterar a chave, exclua LF-Tag e adicione uma com a chave necessária. Além da permissão `Alter`, você também precisa da permissão do IAM `lakeformation:UpdateLFTag` para atualizar os valores.

Quando você exclui um LF-Tag valor, nenhuma verificação é executada para verificar a presença desse LF-Tag valor em qualquer recurso do Catálogo de Dados. Se o LF-Tag valor excluído estiver associado a um recurso, ele não estará mais visível para o recurso, e todos os diretores que receberam permissões nesse par de valores-chave não terão mais as permissões.

Antes de excluir um LF-Tag valor, você pode, opcionalmente, usar o `remove-lf-tags-from-resource`comando de comando para remover os recursos LF-Tag do Catálogo de Dados que tenham o valor que você deseja excluir e, em seguida, remarcar o recurso com os valores que deseja manter.

Somente administradores do data lake, o LF-Tag criador e os diretores que têm `Alter` permissões no LF-Tag podem atualizar um. LF-Tag

Você pode atualizar um LF-Tag usando o AWS Lake Formation console, a API ou o AWS Command Line Interface (AWS CLI).

Console

Para atualizar um LF-Tag (console)

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

   Faça login como administrador, LF-Tag criador ou diretor do data lake com `Alter` permissão no LF-Tag.
2. No painel de navegação, em Permissões LF-Tags e permissões, escolha LF-Tags.
3. Na LF-Tagspágina, selecione um e LF-Tag, em seguida, escolha Editar.
4. Na caixa de LF-Tag diálogo Editar, adicione ou remova LF-Tag valores.

   Para inserir vários Valores, insira uma lista delimitada por vírgulas e pressione Enter ou insira um valor por vez e escolha Adicionar após cada um.
5. Escolha Salvar.

AWS CLI

Para atualizar um LF-Tag (AWS CLI)

- Insira um comando `update-lf-tag`. Forneça um ou ambos os argumentos a seguir:
  - `--tag-values-to-add`
  - `--tag-values-to-delete`

Example

O exemplo a seguir substitui o valor `vp` `vice-president` pelo valor da LF-Tag chave`level`.

```
aws lakeformation update-lf-tag --tag-key level --tag-values-to-add vice-president
--tag-values-to-delete vp
```

Excluindo LF-Tags

Você pode excluir os LF-Tags que não estão mais em uso. Nenhuma verificação é realizada quanto à presença do LF-Tag em um recurso do Catálogo de Dados. Se o excluído LF-Tag estiver associado a um recurso, ele não estará mais visível para o recurso, e todos os diretores que receberam permissões sobre ele LF-Tag não terão mais as permissões.

Antes de excluir um LF-Tag, você pode, opcionalmente, usar o `remove-lf-tags-from-resource`comando para remover o LF-Tag de todos os recursos.

Somente administradores do data lake, o LF-Tag criador ou um diretor que tenha `Drop` permissão no LF-Tag podem excluir um LF-Tag. Além da `Drop` permissão, o diretor também precisa da permissão `lakeformation:DeleteLFTag` do IAM para excluir um LF-Tag.

Você pode excluir um LF-Tag usando o AWS Lake Formation console, a API ou o AWS Command Line Interface (AWS CLI).

Console

Para excluir um LF-Tag (console)

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

   Faça login como administrador de data lake.
2. No painel de navegação, em Permissões LF-Tags e permissões, escolha LF-Tags.
3. Na LF-Tagspágina, selecione um e LF-Tag, em seguida, escolha Excluir.
4. No ambiente Delete tag? caixa de diálogo, para confirmar a exclusão, insira o valor da LF-Tag chave no campo designado e escolha Excluir.

AWS CLI

Para excluir um LF-Tag (AWS CLI)

- Insira um comando `delete-lf-tag`. Forneça a chave do LF-Tag a ser excluído.

  Example

  O exemplo a seguir exclui o LF-Tag com a chave`region`.

  ```
  aws lakeformation delete-lf-tag --tag-key region
  ```

Listagem LF-Tags

Você pode listar aqueles em LF-Tags que você tem as `Associate` permissões `Describe` ou. Os valores listados com cada LF-Tag chave são os valores sobre os quais você tem permissões.

LF-Tag O criador tem permissões implícitas para ver o LF-Tags que ele criou.

Os administradores do data lake podem ver tudo o LF-Tags que está definido na AWS conta local e tudo LF-Tags para o qual as `Associate` permissões `Describe` e foram concedidas à conta local a partir de contas externas. O administrador do data lake pode ver todos os valores de todos LF-Tags.

Você pode listar LF-Tags usando o AWS Lake Formation console, a API ou o AWS Command Line Interface (AWS CLI).

Console

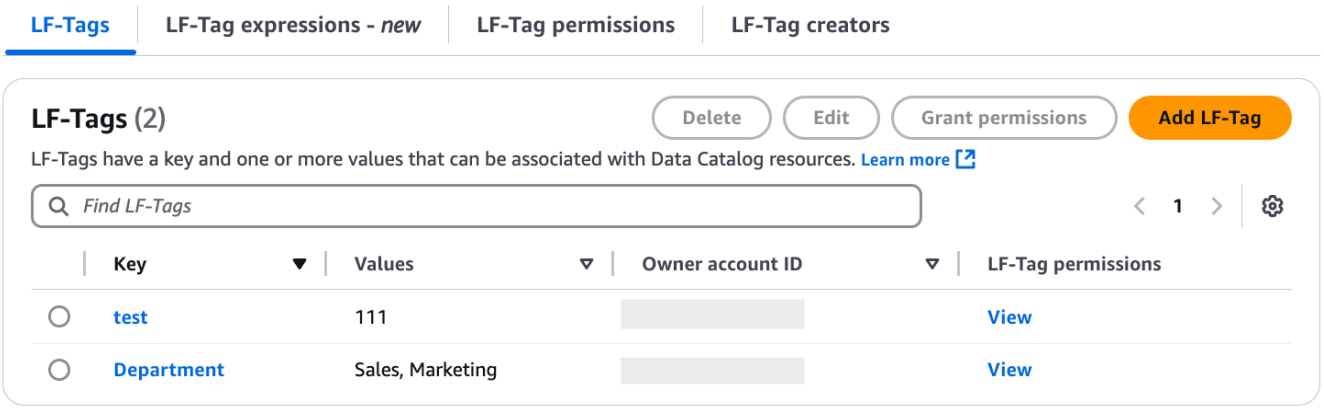
Para listar LF-Tags (console)

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

   Faça login como LF-Tag criador, administrador do data lake ou diretor que recebeu permissões LF-Tags e tem a permissão do `lakeformation:ListLFTags` IAM.

2. No painel de navegação, em Permissões LF-Tags e permissões, escolha LF-Tags.

   A página LF-Tags é exibida.



   Verifique a coluna ID da conta do proprietário para determinar as LF-Tags que foram compartilhadas com sua conta a partir de uma conta externa.

AWS CLI

Para listar LF-Tags (AWS CLI)

- Execute o comando a seguir como administrador do data lake ou como principal que tenha recebido permissões LF-Tags e que tenha a permissão `lakeformation:ListLFTags` do IAM.

```
aws lakeformation list-lf-tags
```

A saída é semelhante à seguinte.

```
{
    "LFTags": [
        {
            "CatalogId": "111122223333",
            "TagKey": "level",
            "TagValues": [
                "director",
                "vp",
                "c-level"
            ]
        },
        {
            "CatalogId": "111122223333",
            "TagKey": "module",
            "TagValues": [
                "Orders",
                "Sales",
                "Customers"
            ]
        }
    ]
}
```

Para ver também LF-Tags se foram concedidos por contas externas, inclua a opção de comando`--resource-share-type ALL`.

```
aws lakeformation list-lf-tags --resource-share-type ALL
```

A saída é semelhante à seguinte. Observe a chave `NextToken`, que indica que há mais para listar.

```
{
    "LFTags": [
        {
            "CatalogId": "111122223333",
            "TagKey": "level",
            "TagValues": [
                "director",
                "vp",
                "c-level"
            ]
        },
        {
            "CatalogId": "111122223333",
            "TagKey": "module",
            "TagValues": [
                "Orders",
                "Sales",
                "Customers"
            ]
        }
    ],
    "NextToken": "eyJleHBpcmF0aW...ZXh0Ijp0cnVlfQ=="
}
```

Repita o comando e adicione o `--next-token` argumento para ver todos os itens locais LF-Tags restantes LF-Tags que foram concedidos por contas externas. LF-Tags das contas externas estão sempre em uma página separada.

```
aws lakeformation list-lf-tags --resource-share-type ALL
--next-token eyJleHBpcmF0aW...ZXh0Ijp0cnVlfQ==
```

```
{
    "LFTags": [
        {
            "CatalogId": "123456789012",
            "TagKey": "region",
            "TagValues": [
```

```
                "central",
                "south"
            ]
        }
    ]
}
```

API

Você pode usar os SDKs disponíveis para o Lake Formation para listar as tags que o solicitante tem permissão para visualizar.

```
import boto3

client = boto3.client('lakeformation')
...

response = client.list_lf_tags(
    CatalogId='string',
    ResourceShareType='ALL',
    MaxResults=50'
)
```

Este comando retorna um objeto `dict` com a seguinte estrutura:

```
{
    'LFTags': [
        {
            'CatalogId': 'string',
            'TagKey': 'string',
            'TagValues': [
                'string',
            ]
        },
    ],
    'NextToken': 'string'
}
```

Para mais informações sobre as permissões necessárias, consulte Referência de personas e permissões do IAM do Lake Formation.

Atribuição LF-Tags aos recursos do Catálogo de Dados

Você pode LF-Tags atribuir recursos do Catálogo de Dados (bancos de dados, tabelas e colunas) para controlar o acesso a esses recursos. Somente os diretores que recebem a correspondência LF-Tags (e os diretores que recebem acesso com o método de recurso nomeado) podem acessar os recursos.

Se uma tabela herdar um LF-Tag de um banco de dados ou uma coluna herdar um LF-Tag de uma tabela, você poderá substituir o valor herdado atribuindo um novo valor à chave. LF-Tag

O número máximo LF-Tags que você pode atribuir a um recurso é 50.

Tópicos

- Requisitos para gerenciar tags atribuídas aos recursos
- LF-Tags Atribuir a uma coluna da tabela
- LF-Tags Atribuir a um recurso do Catálogo de Dados
- Atualização LF-Tags para um recurso
- Removendo LF-Tag de um recurso

Requisitos para gerenciar tags atribuídas aos recursos

Para atribuir um LF-Tag a um recurso do Catálogo de Dados, você deve:

- Tenha a `ASSOCIATE` permissão do Lake Formation no LF-Tag.
- Tenha a permissão `lakeformation:AddLFTagsToResource` do IAM.
- Tenha cola: GetDatabase permissão em um banco de dados do Glue.
- Seja o proprietário do recurso (criador), tenha a permissão do Lake Formation `Super` no recurso com a opção `GRANT` ou tenha as seguintes permissões com a opção `GRANT`:
  - Para bancos de dados na mesma AWS conta: `DESCRIBECREATE_TABLE`,`ALTER`, e `DROP`
  - Para bancos de dados em uma conta externa: `DESCRIBE`, `CREATE_TABLE` e `ALTER`
  - Para tabelas (e colunas): `DESCRIBE`, `ALTER`, `DROP`, `INSERT`, `SELECT` e `DELETE`

Além disso, o LF-Tag e o recurso ao qual ele está sendo atribuído devem estar na mesma AWS conta.

Para remover a LF-Tag de um recurso do catálogo de dados, você deve atender a esses requisitos e também ter a permissão `lakeformation:RemoveLFTagsFromResource` do IAM.

LF-Tags Atribuir a uma coluna da tabela

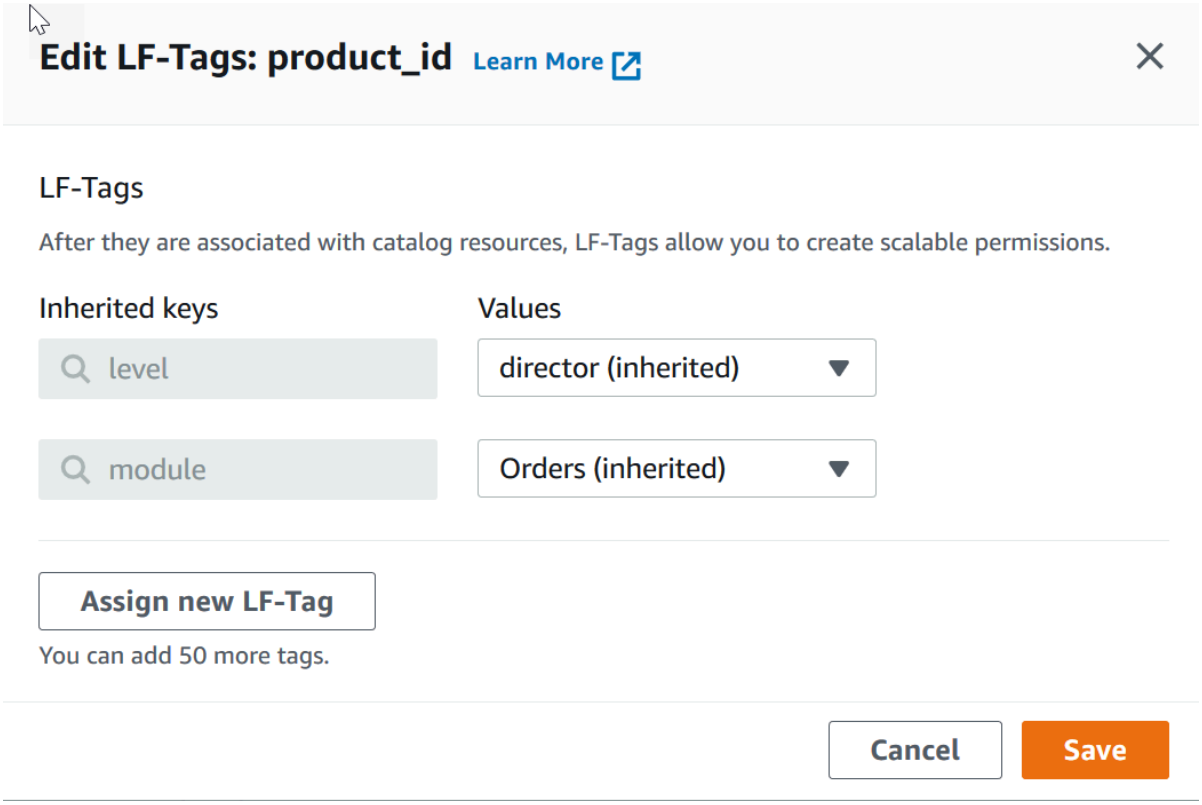
Para LF-Tags atribuir a uma coluna da tabela (console)

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

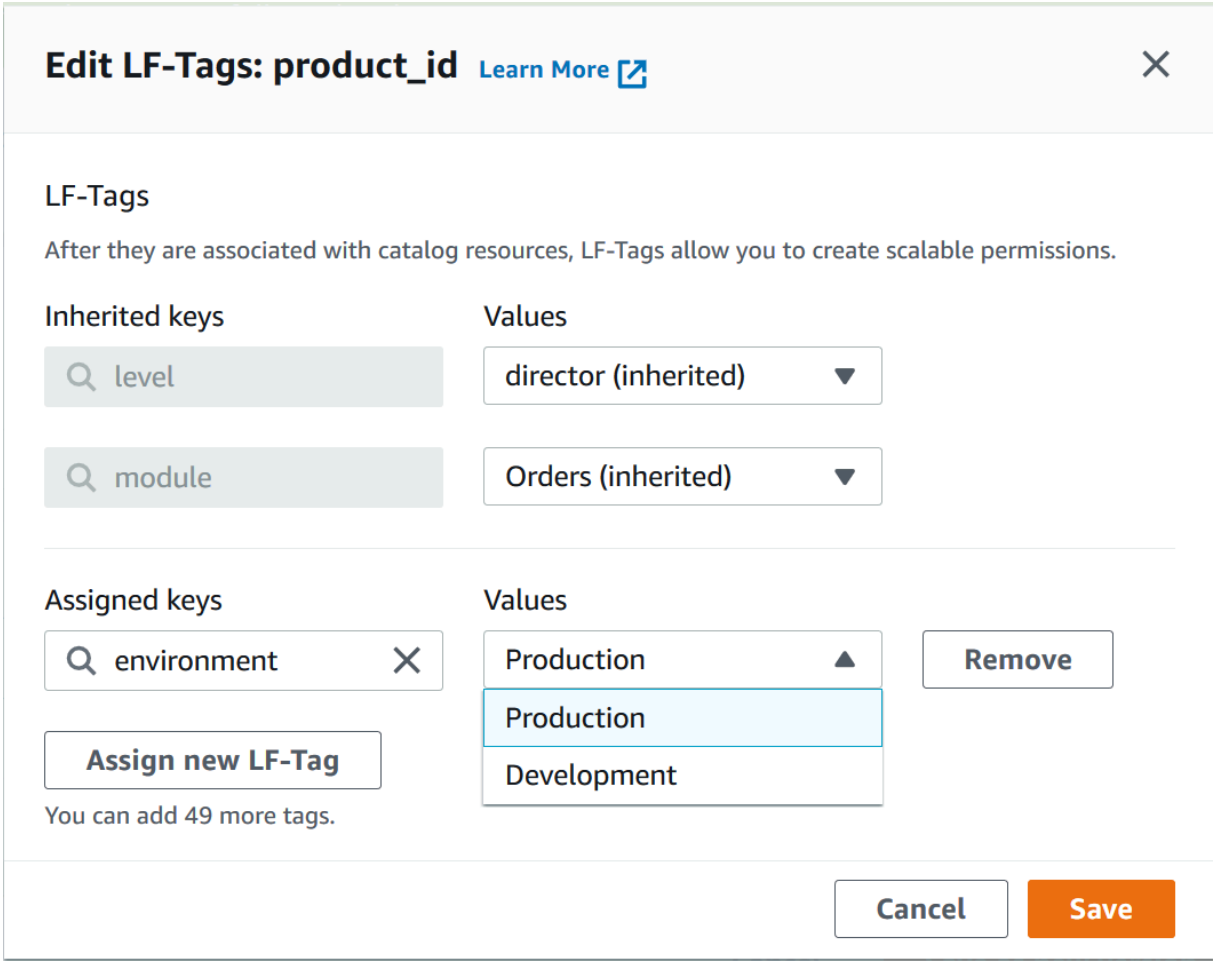
   Faça login como um usuário que atende aos requisitos listados acima.
2. No painel de navegação, selecione Tabelas.
3. Escolha um nome de tabela (não o botão de opção ao lado do nome da tabela).
4. Na página de detalhes da tabela, na seção Esquema, escolha Editar esquema.
5. Na página Editar esquema, selecione uma ou mais colunas e escolha Editar LF-Tags.

> **Note**
>
> Se você pretende adicionar ou excluir colunas e salvar uma nova versão, faça isso primeiro. Em seguida, edite LF-Tags o.

A caixa LF-Tags de diálogo Editar é exibida e exibe todas as LF-Tags que são herdadas da tabela.



6. (Opcional) Na lista Valores ao lado do campo Chaves herdadas, escolha um valor para substituir o valor herdado.
7. (Opcional) Escolha Atribuir novo LF-Tag. Em seguida, em Chaves atribuídas, escolha uma chave e, em Valores, escolha um valor para a chave.



8. (Opcional) Escolha Atribuir novo LF-Tag novamente para adicionar outro LF-Tag.
9. Escolha Salvar.

LF-Tags Atribuir a um recurso do Catálogo de Dados

Console

Para atribuir LF-Tags a um banco de dados ou tabela do Catálogo de Dados

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

   Faça login como um usuário que atende aos requisitos listados anteriormente.

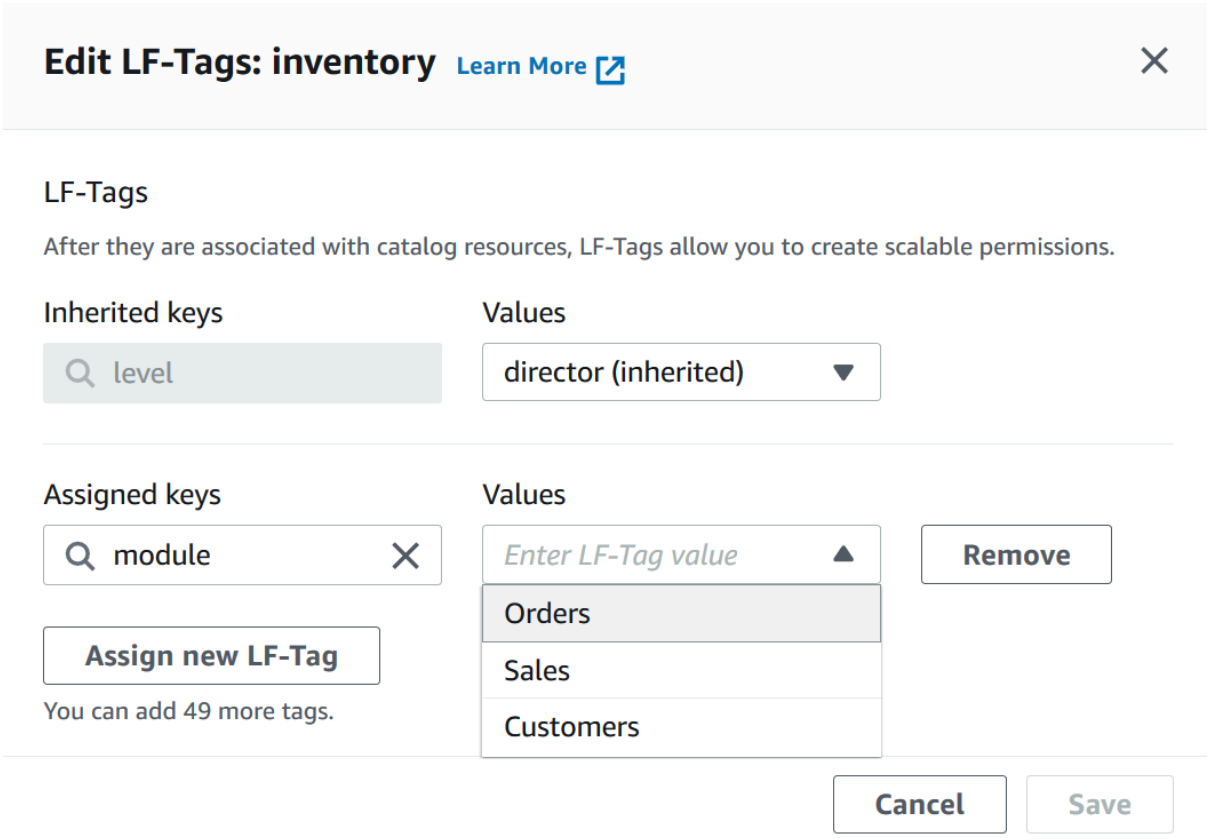
2. No painel de navegação, em catálogo de dados, faça o seguinte:

   - Para atribuir LF-Tags aos bancos de dados, escolha Bancos de dados.
   - Para atribuir LF-Tags às tabelas, escolha Tabelas.

3. Escolha um banco de dados ou tabela e, no menu Ações, escolha Editar LF-Tags.

   A caixa de *`resource-name`* diálogo Editar LF-Tags: é exibida.

   Se uma tabela herdar do banco LF-Tags de dados que a contém, a janela exibirá o herdado LF-Tags. Caso contrário, ele exibirá o texto “Não há herdados LF-Tags associados ao recurso”.



4. (Opcional) Se uma tabela tiver sido herdada LF-Tags, na lista Valores ao lado do campo Chaves herdadas, você poderá escolher um valor para substituir o valor herdado.

5. Para atribuir um novo LF-Tags, execute estas etapas:

   a. Escolha Atribuir novo LF-Tag.

   b. No campo Chaves atribuídas, escolha uma LF-Tag chave e, no campo Valores, escolha um valor.

c. (Opcional) Escolha Atribuir novo LF-Tag novamente para atribuir um adicional LF-Tag.

6. Escolha Salvar.

AWS CLI

Para atribuir LF-Tags a um recurso do Catálogo de Dados

- Execute o comando `add-lf-tags-to-resource`.

  O exemplo a seguir atribui o LF-Tag `module=orders` à tabela `orders` no banco de dados`erp`. Ele usa a sintaxe de atalho para o argumento `--lf-tags`. A propriedade `CatalogID` para `--lf-tags` é opcional. Se não for fornecido, o ID do catálogo do recurso (nesse caso, a tabela) será assumido.

```
aws lakeformation add-lf-tags-to-resource --resource '{ "Table":
 {"DatabaseName":"erp", "Name":"orders"}}' --lf-tags
 CatalogId=111122223333,TagKey=module,TagValues=orders
```

  A seguir está o resultado se o comando for bem-sucedido.

```
{
    "Failures": []
}
```

  O próximo exemplo atribui dois LF-Tags à `sales` tabela e usa a sintaxe JSON para o argumento. `--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"]}]'
```

  O próximo exemplo atribui o LF-Tag `level=director` à `total` coluna da tabela`sales`.

```
aws lakeformation add-lf-tags-to-resource --resource '{ "TableWithColumns":
 {"DatabaseName":"erp", "Name":"sales", "ColumnNames":["total"]}}' --lf-tags
 TagKey=level,TagValues=director
```

Atualização LF-Tags para um recurso

Para atualizar um LF-Tag para um recurso do Catálogo de Dados (AWS CLI)

- Use o comando `add-lf-tags-to-resource`, conforme descrito no procedimento anterior.

  Adicionar um LF-Tag com a mesma chave de um existente LF-Tag, mas com um valor diferente, atualiza o valor existente.

Removendo LF-Tag de um recurso

Para remover um LF-Tag para um recurso do Catálogo de Dados (AWS CLI)

- Execute o comando `remove-lf-tags-from-resource`.

  Se uma tabela tiver um LF-Tag valor que substitua o valor herdado do banco de dados pai, removê-lo LF-Tag da tabela restaurará o valor herdado. Esse comportamento também se aplica a uma coluna que substitui os valores-chave herdados da tabela.

  O exemplo a seguir LF-Tag `level=director` remove o da `total` coluna da `sales` tabela. A propriedade `CatalogID` para `--lf-tags` é opcional. Se não for fornecido, o ID do catálogo do recurso (nesse caso, a tabela) será assumido.

```
aws lakeformation remove-lf-tags-from-resource
--resource ' { "TableWithColumns":
{ "DatabaseName": "erp",  "Name": "sales",  "ColumnNames":[ "total"]}}'
--lf-tags  CatalogId=111122223333,TagKey=level,TagValues=director
```

Visualização LF-Tags atribuída a um recurso

Você pode visualizar os LF-Tags que estão atribuídos a um recurso do Catálogo de Dados. Você deve ter a `ASSOCIATE` permissão `DESCRIBE` ou em um LF-Tag para visualizá-lo.

Console

Para ver os LF-Tags que estão atribuídos a um recurso (console)

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

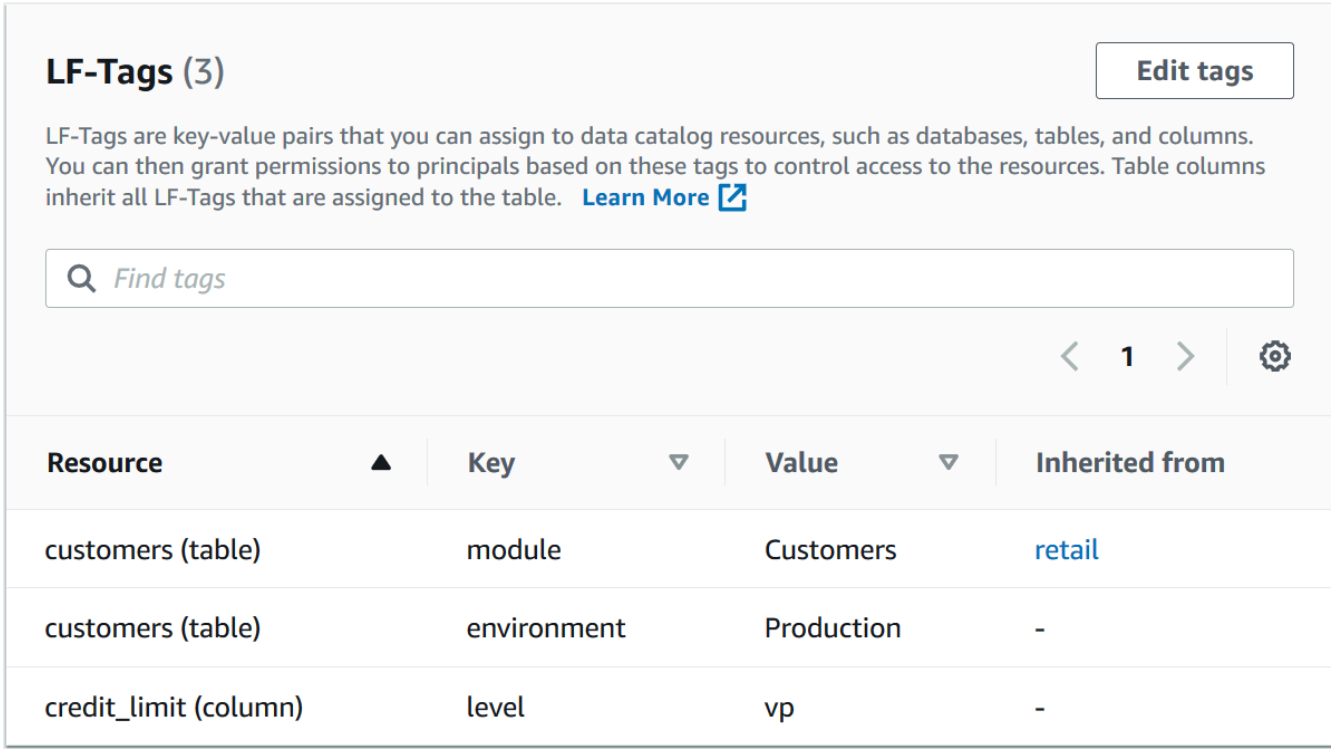
Faça login como administrador do data lake, proprietário do recurso ou usuário que tenha recebido permissões do Lake Formation no recurso.

2. No painel de navegação, no título catálogo de dados, faça o seguinte:

   - Para ver a exibição LF-Tags atribuída a um banco de dados, escolha Bancos de dados.
   - Para ver a exibição LF-Tags atribuída a uma tabela, escolha Tabelas.

3. Na página Tabelas ou Bancos de Dados, escolha o nome do banco de dados ou da tabela. Em seguida, na página de detalhes, desça até a LF-Tagsseção.

   A captura de tela a seguir mostra o LF-Tags atribuído a uma `customers` tabela, que está contida no `retail` banco de dados. O `module` LF-Tag é herdado do banco de dados. A `credit_limit` coluna tem o `level=vp` LF-Tag atribuído.



AWS CLI

Para ver os LF-Tags que estão atribuídos a um recurso (AWS CLI)

- Digite um comando semelhante ao seguinte:

```
aws lakeformation get-resource-lf-tags --show-assigned-lf-tags --
resource '{ "Table": {"CatalogId":"111122223333", "DatabaseName":"erp",
 "Name":"sales"}}'
```

O comando retorna a seguinte saída.

```
{
    "TableTags": [
        {
            "CatalogId": "111122223333",
            "TagKey": "module",
            "TagValues": [
                "sales"
            ]
        },
        {
            "CatalogId": "111122223333",
            "TagKey": "environment",
            "TagValues": [
                "development"
            ]
        }
    ],
    "ColumnTags": [
        {
            "Name": "total",
            "Tags": [
                {
                    "CatalogId": "111122223333",
                    "TagKey": "level",
                    "TagValues": [
                        "director"
                    ]
                }
            ]
        }
    ]
}
```

Essa saída mostra somente LF-Tags aquelas atribuídas explicitamente, não herdadas. Se você quiser ver tudo LF-Tags em todas as colunas, inclusive as herdadas LF-Tags, omita a `--show-assigned-lf-tags` opção.

Visualizando os recursos aos quais um LF-Tag está atribuído

Você pode visualizar todos os recursos do Catálogo de Dados aos quais uma LF-Tag chave específica está atribuída. Para fazer isso, você precisa das seguintes permissões do Lake Formation:

- `Describe`ou `Associate` no LF-Tag.
- `Describe` ou qualquer outra permissão do Lake Formation sobre o recurso.

Além disso, você precisa das seguintes permissões AWS Identity and Access Management (IAM):

- `lakeformation:SearchDatabasesByLFTags`
- `lakeformation:SearchTablesByLFTags`

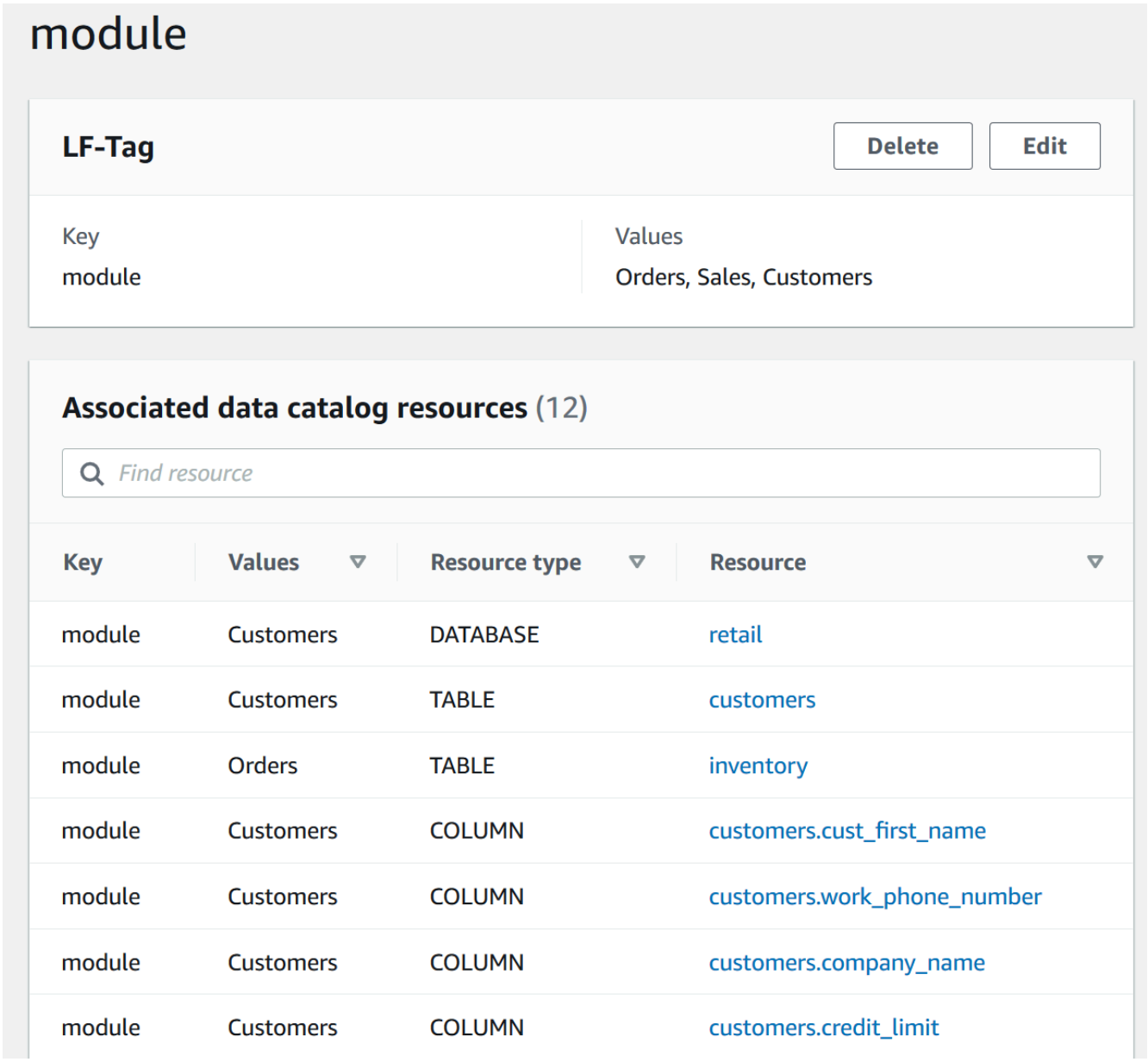
Console

Para visualizar os recursos aos quais a LF-Tag está atribuído (console)

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

   Faça login como administrador do data lake ou como usuário que atenda aos requisitos listados anteriormente.
2. No painel de navegação, em Permissões LF-Tags e permissões, escolha LF-Tags.
3. Escolha uma LF-Tag chave (não o botão de opção ao lado do nome da chave).

   A página de LF-Tag detalhes exibe uma lista dos recursos aos quais LF-Tag foram atribuídos.



AWS CLI

Para visualizar os recursos aos quais um LF-Tag está atribuído

- Execute um comando `search-tables-by-lf-tags` or `search-databases-by-lf-tags`.

Example

O exemplo a seguir lista tabelas e colunas que têm o `level=vp` LF-Tag atribuído. Para cada tabela e coluna listada, todas as atribuídas LF-Tags à tabela ou coluna são geradas, não apenas a expressão de pesquisa.

```
aws lakeformation search-tables-by-lf-tags --expression
 TagKey=level,TagValues=vp
```

Para mais informações sobre as permissões necessárias, consulte Referência de personas e permissões do IAM do Lake Formation.

Ciclo de vida de um LF-Tag

1. O LF-Tag criador Michael cria um LF-Tag`module=Customers`.
2. Michael `Associate` concede o LF-Tag ao engenheiro de dados Eduardo. Conceder `Associate` concede implicitamente `Describe`.
3. Michael concede `Super` na mesa `Custs` a Eduardo com a opção de concessão, para que Eduardo possa LF-Tags se inscrever na mesa. Para obter mais informações, consulte Atribuição LF-Tags aos recursos do Catálogo de Dados.
4. Eduardo atribui o LF-Tag `module=customers` à mesa. `Custs`
5. Michael faz a seguinte concessão à engenheira de dados Sandra (em pseudocódigo).

```
GRANT (SELECT, INSERT ON TABLES) ON TAGS module=customers TO Sandra WITH GRANT OPTION
```

6. Sandra faz a seguinte concessão à analista de dados Maria.

```
GRANT (SELECT ON TABLES) ON TAGS module=customers TO Maria
```

Agora, Maria pode executar consultas na tabela `Custs`.

Consulte também

- Controle de acesso a metadados

Comparação do controle de acesso baseado em tags do Lake Formation com o controle de acesso baseado em atributos do IAM

Attribute-based controle de acesso (ABAC) é uma estratégia de autorização que define permissões com base em atributos. Em AWS, esses atributos são chamados de tags. Você pode anexar tags aos recursos do IAM, incluindo entidades do IAM (usuários ou funções) e aos AWS recursos. É possível criar uma única política de ABAC ou um pequeno conjunto de políticas para suas entidades do IAM. Essas políticas de ABAC podem ser criadas para permitir operações quando a tag da entidade principal corresponder à tag de recurso. O ABAC é útil em ambientes que estão crescendo rapidamente e ajuda em situações em que o gerenciamento de políticas se torna um problema.

As equipes de segurança e governança da nuvem usam o IAM para definir políticas de acesso e permissões de segurança para todos os recursos, incluindo buckets do Amazon S3, instâncias do Amazon EC2 e quaisquer recursos que você possa referenciar com um ARN. As políticas do IAM definem permissões amplas (grosseiras) para seus recursos de data lake, por exemplo, para permitir ou negar acesso no nível de bucket, prefixo ou banco de dados do Amazon S3. Para obter mais informações sobre o IAM ABAC, consulte Para que serve o ABAC? AWS no Guia do usuário do IAM.

Por exemplo, é possível criar três funções com a chave de tag `project-access`. Defina o valor da tag da primeira função como `Dev`, a segunda como `Marketing` e a terceira como `Support`. Atribua tags com o valor apropriado aos recursos. Depois disso, é possível usar uma única política que permitirá o acesso quando a função e o recurso estiverem marcados com o mesmo valor para `project-access`.

As equipes de governança de dados usam o Lake Formation para definir permissões refinadas para recursos específicos do data lake. LF-Tags são atribuídos aos recursos do Catálogo de Dados (bancos de dados, tabelas e colunas) e são concedidos aos diretores. Um diretor com LF-Tags a mesma proporção LF-Tags de um recurso pode acessar esse recurso. As permissões do Lake Formation são secundárias às permissões do IAM. Por exemplo, se as permissões do IAM não permitirem que um usuário acesse um data lake, o Lake Formation não concederá acesso a nenhum recurso dentro desse data lake para esse usuário, mesmo que o principal e o recurso tenham correspondência LF-Tags.

O controle de acesso baseado em tags do Lake Formation (LF-TBAC) funciona com o IAM ABAC para fornecer níveis adicionais de permissões para seus dados e recursos do Lake Formation.

- As permissões de TBAC do Lake Formation são dimensionadas com inovação. Não é mais necessário que um administrador atualize as políticas existentes para permitir o acesso a novos recursos. Por exemplo, suponha que você use uma estratégia ABAC do IAM com a tag `project-`

`access` para fornecer acesso a bancos de dados específicos dentro do Lake Formation. Usando LF-TBAC, o LF-Tag `Project=SuperApp` é atribuído a tabelas ou colunas específicas, e o mesmo LF-Tag é concedido a um desenvolvedor desse projeto. Por meio do IAM, o desenvolvedor pode acessar o banco de dados, e LF-TBAC as permissões concedem ao desenvolvedor acesso adicional a tabelas ou colunas específicas dentro das tabelas. Se uma nova tabela for adicionada ao projeto, o administrador do Lake Formation só precisará atribuir a tag à nova tabela para que o desenvolvedor tenha acesso a ela.

- O Lake Formation TBAC exige menos políticas de IAM. Como você usa as políticas do IAM para conceder acesso de alto nível aos recursos do Lake Formation e o TBAC do Lake Formation para gerenciar um acesso mais preciso aos dados, você cria menos políticas do IAM.
- Ao usar o Lake Formation TBAC, as equipes podem mudar e crescer rapidamente. Isso ocorre porque as permissões para novos recursos são concedidas automaticamente com base em atributos. Por exemplo, se um novo desenvolvedor ingressar no projeto, é fácil conceder acesso a esse desenvolvedor associando a função do IAM ao usuário e, em seguida, atribuindo o necessário LF-Tags ao usuário. Você não precisa alterar a política do IAM para apoiar um novo projeto ou criar um novo LF-Tags.
- Finer-grained as permissões são possíveis usando o Lake Formation TBAC. As políticas do IAM concedem acesso aos recursos de nível superior, como bancos de dados ou tabelas do catálogo de dados. Ao usar o Lake Formation TBAC, você pode conceder acesso a tabelas ou colunas específicas que contêm valores de dados específicos.

> **Note**
>
> As tags IAM não são as mesmas que LF-Tags. Essas etiquetas não são intercambiáveis. LF-Tags são usadas para conceder permissões do Lake Formation e as tags do IAM são usadas para definir políticas do IAM.

Gerenciando LF-Tag expressões para controle de acesso a metadados

LF-Tag expressões são expressões lógicas compostas por um ou mais LF-Tags (pares de valores-chave) usadas para conceder permissões em AWS Glue Data Catalog recursos. LF-Tag expressões permitem que você defina regras que controlam o acesso aos seus recursos de dados com base em suas tags de metadados. É possível salvar essas expressões e reutilizá-las em várias concessões de permissão, garantindo a consistência e facilitando o gerenciamento de alterações na ontologia de tags ao longo do tempo.

Em uma determinada LF-Tag expressão, as chaves de tag são combinadas usando a operação AND, enquanto os valores são combinados usando a operação OR. Por exemplo, a expressão de tag `content_type:Sales AND location:US` representa recursos relacionados aos dados de vendas nos EUA.

Você pode criar até 1000 LF-Tag expressões em um Conta da AWS. Essas expressões proporcionam uma forma flexível e escalável de gerenciar permissões com base em tags de metadados, garantindo que somente usuários ou aplicações autorizados possam acessar recursos de dados específicos com base nas regras de tag definidas.

LF-Tag as expressões oferecem os seguintes benefícios:

- Reutilização — Ao definir e salvar LF-Tag expressões, você não precisa mais replicar manualmente as mesmas expressões ao atribuir permissões a outros recursos ou entidades principais.
- Consistência — A reutilização de LF-Tag expressões em várias concessões de permissão garante a consistência na forma como as permissões são concedidas e gerenciadas.
- Gerenciamento de ontologia de tags — LF-Tag as expressões ajudam a gerenciar as alterações na ontologia de tags ao longo do tempo, pois você pode atualizar as expressões salvas em vez de modificar as permissões individuais.

Para acessar mais informações sobre o controle de acesso baseado em tags, consulte Controle de acesso baseado em tags do Lake Formation.

LF-Tag criadores de expressões

LF-Tag O criador de expressões é um diretor que tem permissões para criar e gerenciar LF-Tag expressões. Os administradores do Data Lake podem adicionar criadores de LF-Tag expressões usando o console, a CLI, a API ou o SDK do Lake Formation. LF-Tag os criadores de expressões têm permissões implícitas do Lake Formation para criar, atualizar e excluir LF-Tag expressões e para conceder permissões de LF-Tag expressão a outros diretores.

LF-Tag criadores de expressões que não são administradores de data lake recebem `Grant with LF-Tag expression` permissões implícitas `Alter`,`Drop`,`Describe`, e somente para expressões que eles criaram.

Os administradores do data lake também podem conceder permissões LF-Tag concedidas aos criadores de expressões. `Create LF-Tag expression` Em seguida, o criador da LF-Tag expressão pode conceder a permissão para criar LF-Tag expressões para outros diretores.

Tópicos

- Permissões do IAM necessárias para criar LF-Tag expressões
- Adicione criadores de LF-Tag expressões
- Criação de LF-Tag expressões
- Atualizando LF-Tag expressões
- Excluindo expressões LF-Tag
- Listando LF-Tag expressões

Consulte também

- Gerenciando permissões de LF-Tag valor
- Concedendo permissões de data lake usando o método LF-TBAC
- Controle de acesso baseado em tags do Lake Formation

Permissões do IAM necessárias para criar LF-Tag expressões

Você deve configurar permissões para permitir que um diretor do Lake Formation crie LF-Tag expressões. Adicione a seguinte declaração à política de permissões do diretor que precisa ser um criador de LF-Tag expressões.

Note

Embora os administradores do data lake tenham permissões implícitas do Lake Formation para criar, atualizar LF-Tags e excluir LF-Tag expressões, LF-Tags atribuir recursos e conceder LF-Tags e LF-Tag expressar permissão aos diretores, os administradores do data lake também precisam das seguintes permissões do IAM.

Para obter mais informações, consulte Referência de personas e permissões do IAM do Lake Formation.

```
{
"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",
        "lakeformation:CreateLFTagExpression",
        "lakeformation:DeleteLFTagExpression",
        "lakeformation:UpdateLFTagExpression",
        "lakeformation:GetLFTagExpression",
        "lakeformation:ListLFTagExpressions",
        "lakeformation:GrantPermissions",
        "lakeformation:RevokePermissions",
        "lakeformation:BatchGrantPermissions",
        "lakeformation:BatchRevokePermissions"
     ]
}
```

Adicione criadores de LF-Tag expressões

LF-Tag os criadores de expressões podem criar e salvar LF-Tag expressões reutilizáveis, atualizar a chave e os valores da tag, excluir expressões e conceder permissões nos recursos do Catálogo de Dados aos diretores usando o método. LF-TBAC O criador da LF-Tag expressão também pode conceder essas permissões aos diretores.

Você pode criar funções de criador de LF-Tag expressões usando o AWS Lake Formation console, a API ou o AWS Command Line Interface (AWS CLI).

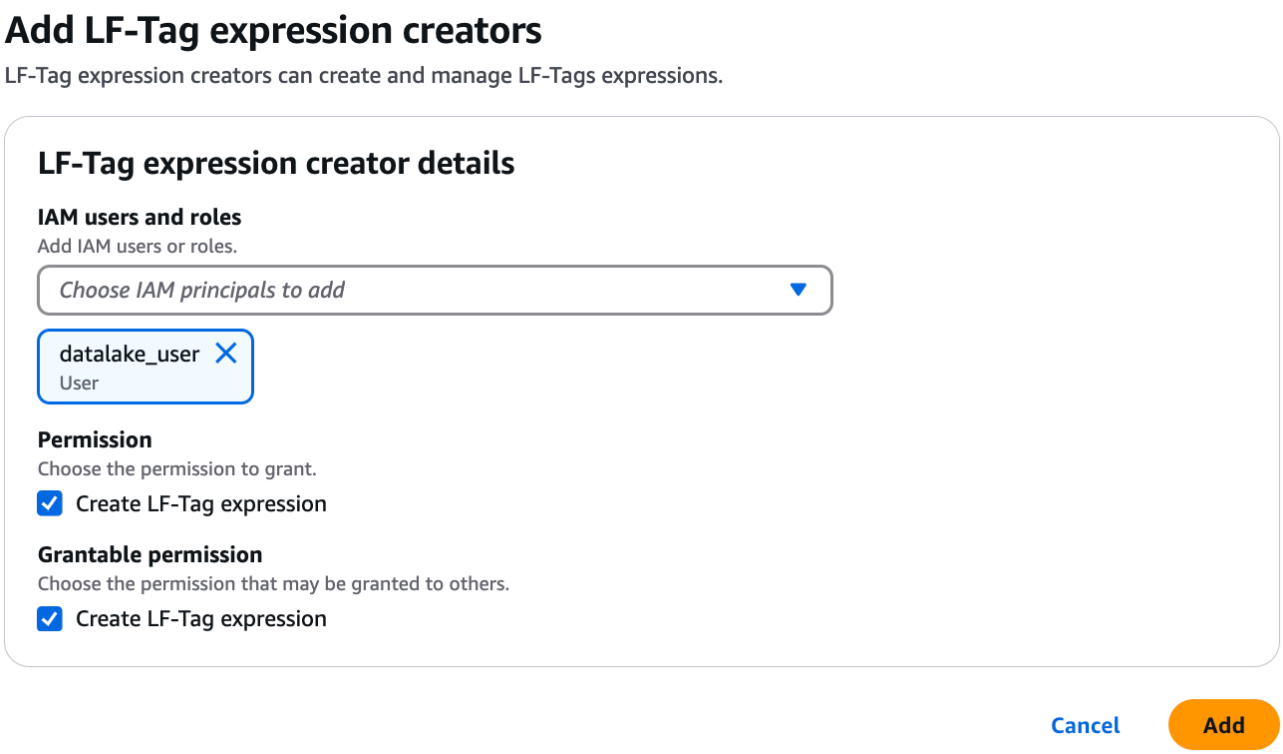
console

Para adicionar um criador de LF-Tag expressão

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

   Faça login como administrador de data lake.
2. No painel de navegação, em Permissões, escolha LF-Tags e permissões.

3. Escolha a guia de LF-Tag expressões.
4. Na seção Criadores de LF-Tag expressões, escolha Adicionar criadores de LF-Tag expressões.



5. Na página Adicionar criadores de LF-Tag expressões, escolha uma função ou usuário do IAM que tenha as permissões necessárias para criar LF-Tag expressões.
6. Marque a caixa de seleção de permissão `Create LF-Tag expression`.
7. (Opcional) Para permitir que as entidades principais selecionadas concedam a permissão `Create LF-Tag expression` às entidades principais, escolha a permissão `Create LF-Tag expression` concedível.
8. Escolha Adicionar.

AWS CLI

```
aws lakeformation grant-permissions --cli-input-json file://grantCreate
{
    "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:user/tag-manager"
    },
    "Resource": {
```

```
        "Catalog": {}
    },
    "Permissions": [
        "CreateLFTagExpression"
    ],
    "PermissionsWithGrantOption": [
        "CreateLFTagExpression"
    ]
}
```

A função de criador de LF-Tag expressões tem a capacidade de criar, atualizar ou excluir LF-Tag expressões.

Permissão	Description
`Create`	Um diretor com essa permissão pode adicionar LF-Tag expressões no data lake.
`Drop`	Um diretor com essa permissão em uma LF-Tag expressão pode excluir uma LF-Tag expressão do data lake.
`Alter`	Um diretor com essa permissão em uma LF-Tag expressão pode atualizar o corpo da LF-Tag expressão.
`Describe`	Um diretor com essa permissão em uma LF-Tag expressão pode visualizar o conteúdo de uma LF-Tag expressão.
`Grant with LF-Tag expression`	Com essa permissão, o destinatário usa a expressão de tag como recurso ao conceder permissões de acesso a dados ou metadados . Conceder `Grant with LF-Tag expression` concede implicita mente `Describe`.
`Super`	Para LF-Tag expressões, a `Super` permissão concede a capacidade de`Describe`, `AlterDrop`, e concede permissões na expressão da tag a outros diretores.

Essas permissões são concedidas. Uma entidade principal que tenha recebido essas permissões com a opção de concessão pode concedê-las a outras entidades principais.

Criação de LF-Tag expressões

Você precisa definir tudo LF-Tags em Lake Formation e atribuí-los aos recursos do Catálogo de Dados antes que eles possam ser usados para criar expressões. Uma LF-Tag expressão consiste em mais uma chave e um ou mais valores possíveis para cada chave.

Depois que o administrador do data lake tiver configurado as permissões necessárias do IAM e as permissões do Lake Formation para a função de criador da LF-Tag expressão, o diretor poderá criar expressões reutilizáveis LF-Tag . O criador da LF-Tag expressão obtém permissões implícitas para atualizar o corpo da expressão e excluir a LF-Tag expressão.

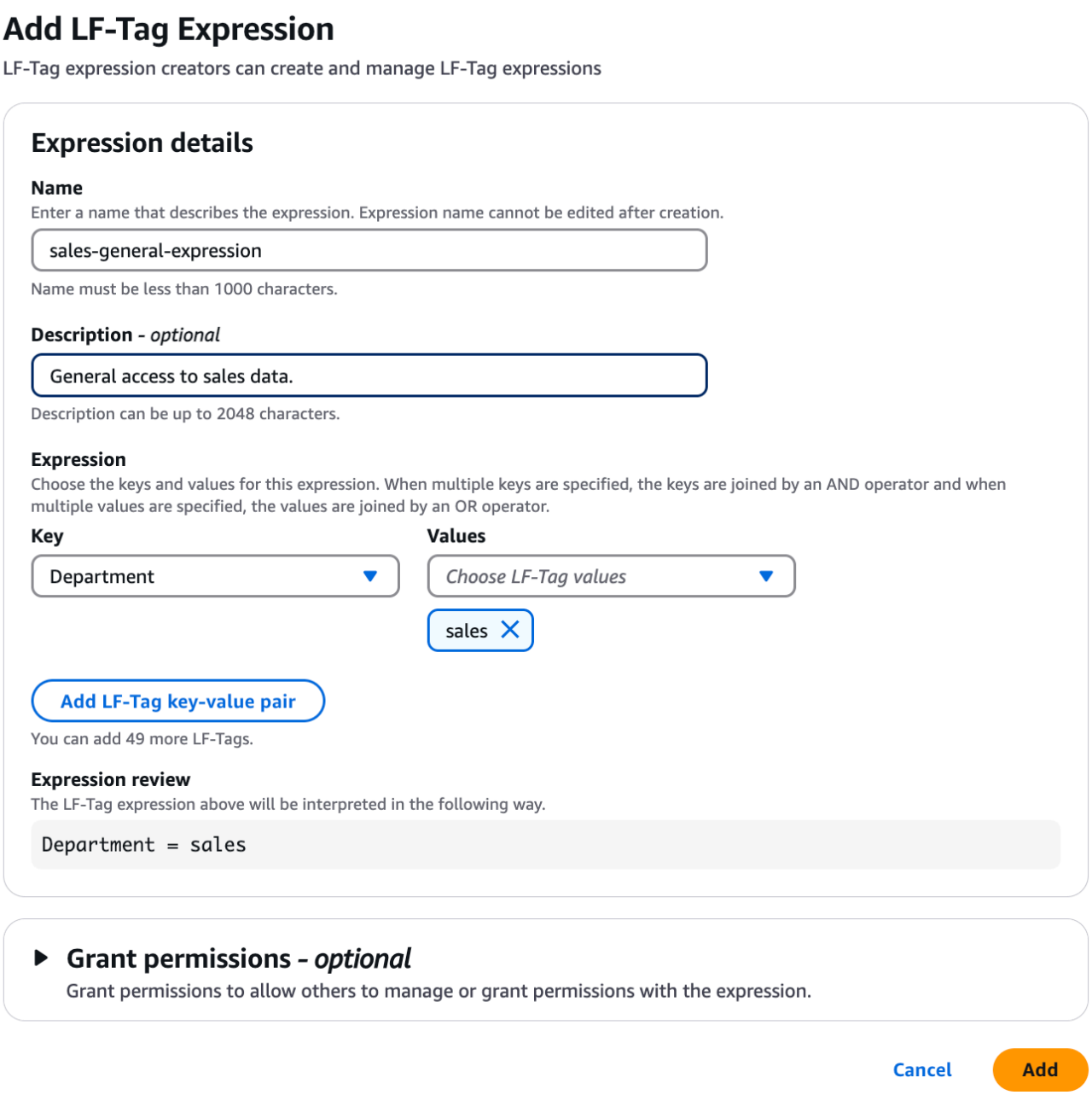
Você pode criar LF-Tag expressões usando o AWS Lake Formation console, a API ou o AWS Command Line Interface (AWS CLI).

Console

Para criar uma LF-Tag expressão

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

   Faça login como diretor com permissões de criador de LF-Tag expressões ou como administrador do data lake.
2. No painel de navegação, em Permissões, escolha LF-Tags e permissões.
3. Escolha LF-Tag expressões. A página Adicionar LF-Tag expressões é exibida.



4. Insira as seguintes informações:

   - Nome: insira um nome exclusivo para a expressão. Não é possível atualizar o nome da expressão.
   - Descrição: forneça uma descrição opcional com os detalhes da expressão.
   - Expressão: crie a expressão especificando as chaves de tag e seus valores associados. É possível adicionar até cinquenta chaves por expressão. É necessário ter permissão `Grant with LF-Tags` do Lake Formation em todas as tags no corpo da expressão.

Cada chave deve ter pelo menos um valor. Para inserir vários valores, insira uma lista delimitada por vírgulas e pressione Enter ou insira um valor por vez e escolha Adicionar após cada um. O número máximo de valores permitidos por chave é mil.

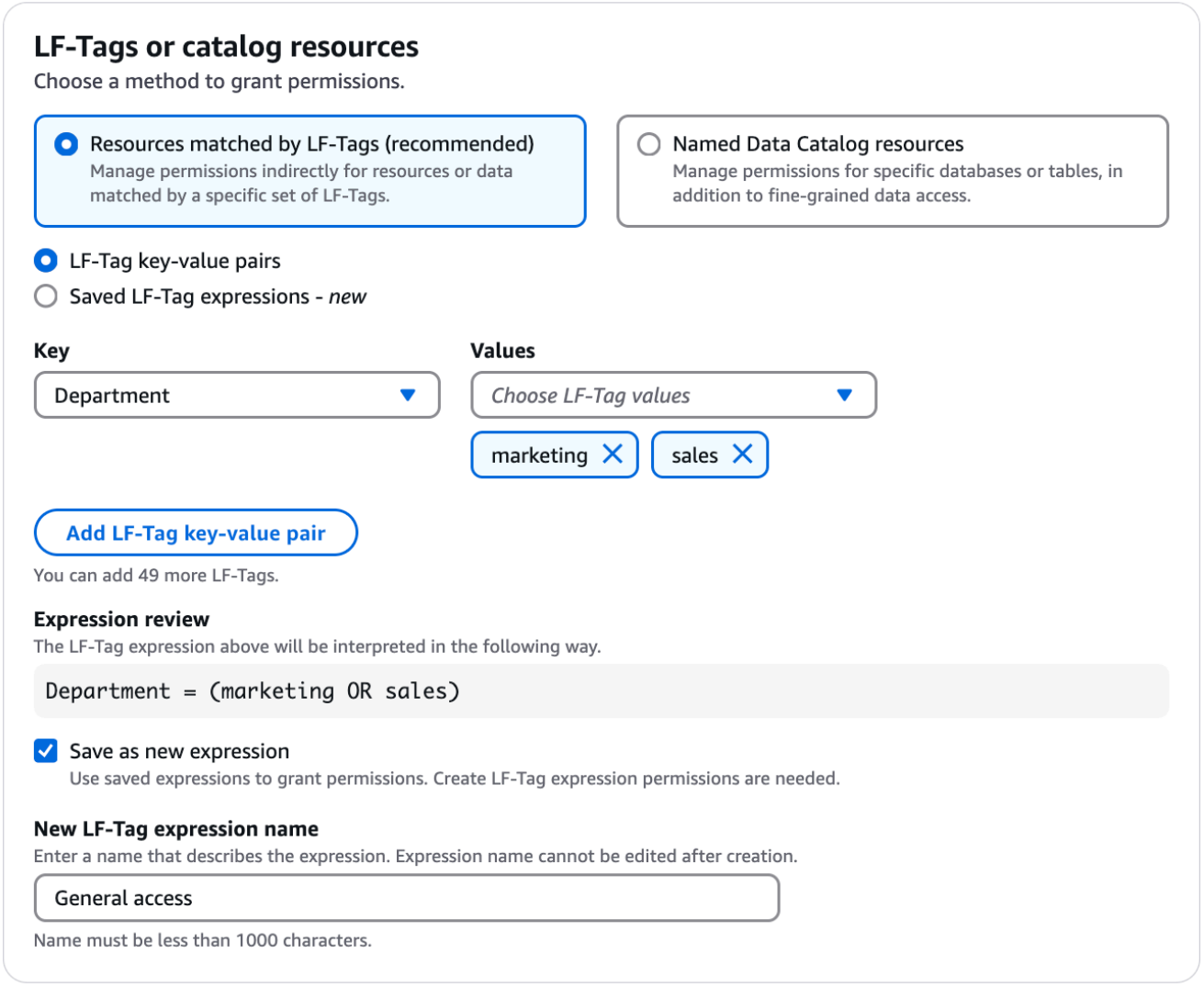
Lake Formation usa a AND/OR lógica para combinar várias chaves e valores em uma expressão. Em um único par (chave: lista de valores), os valores são combinados usando o operador lógico OR. Por exemplo, se o par for (Department : [Sales, Marketing]), significa que a tag vai coincidir se o recurso tiver a tag Department com o valor Sales OU Marketing.

Quando você especifica várias chaves, elas são unidas por um operador lógico AND. Portanto, se a expressão completa for (Department : [Sales, Marketing]) E (Location : [US, Canada]), ela coincidirá com os recursos que têm a tag Department com o valor Sales OU Marketing e também têm a tag Location com o valor US OU Canada. Veja a seguir outro exemplo com várias chaves e valores:

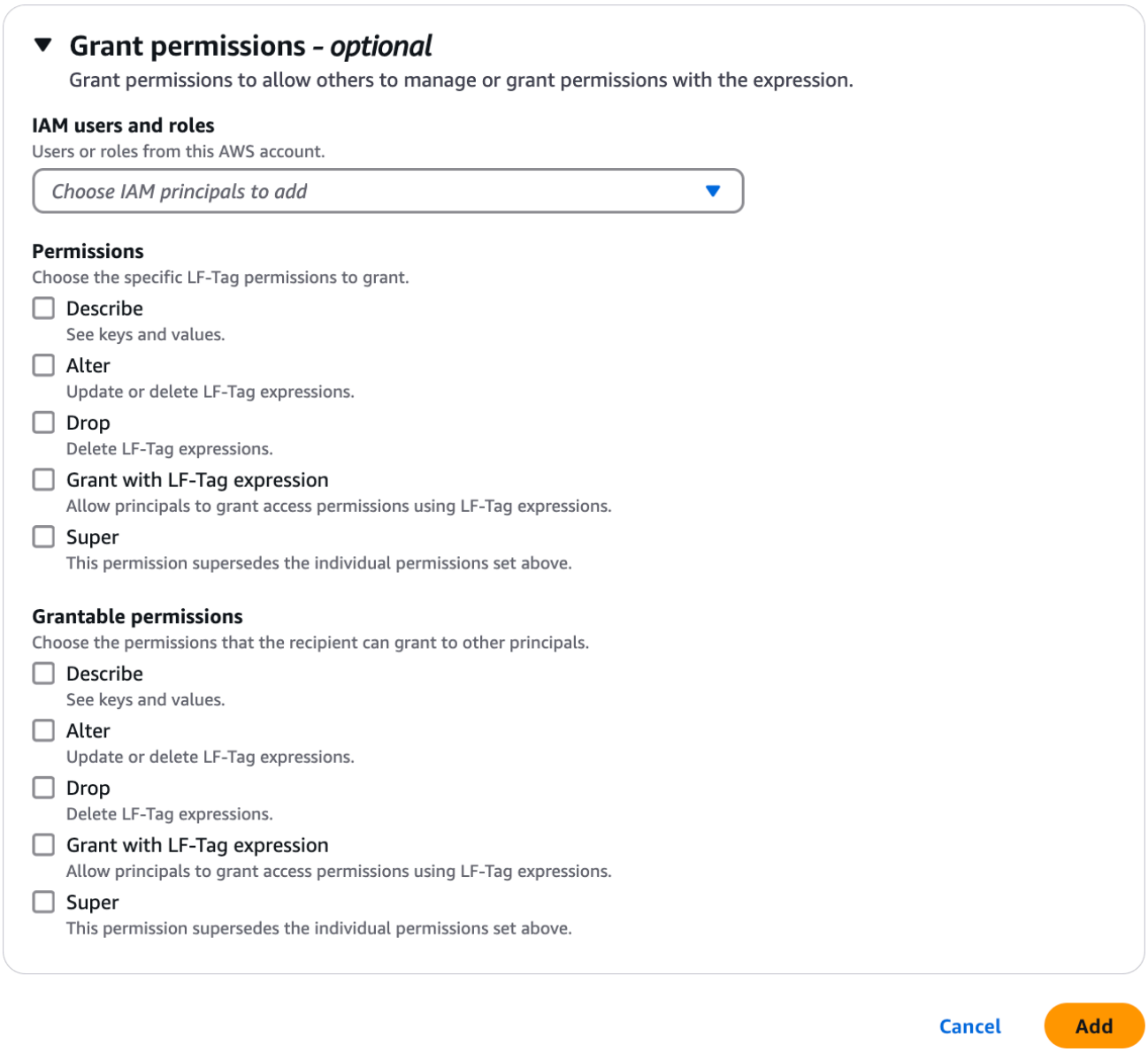
LF-Tag expressão: (ContentType : [Vídeo, Áudio]) AND (Região: [Europa, Ásia]) AND (Departamento: [Engenharia, ProductManagement]).

Essa expressão corresponderia a recursos que têm: - A ContentType tag com valor Vídeo OU Áudio E - A tag Região com valor Europa OU Ásia E - A tag Departamento com valor Engenharia OR ProductManagement.

Você também pode salvar uma expressão de tag ao conceder permissões de data lake usando o. LF-Tags Escolha os pares de chave e valor e selecione a opção Salvar como nova expressão. Insira um nome que descreva a expressão.



5. (Opcional) Em seguida users/roles, escolha a e as permissões na expressão que você deseja conceder a eles na conta. Você também pode escolher permissões para concessão que permitam que os usuários concedam essas permissões a outros usuários na conta. Não é possível conceder permissões entre contas nas expressões de tag.



6. Escolha Adicionar.

AWS CLI

Para criar uma LF-Tag expressão

- Insira um comando `create-lf-tag-expression`.

  O exemplo a seguir cria uma LF-Tag expressão com a tag `Department` com valores `Sales` e `Marketing` AND a tag `Location` com o valorUS.

```
aws lakeformation create-lf-tag-expression \
```

```
-- name "my-tag-expression" \
-- catalog-id "123456789012" \
-- expression '{"Expression":[{"TagKey":"Department","TagValues":
["Sales","Marketing"]},{"TagKey":"Location","TagValues":["US"]}]}'
```

Esse comando da CLI cria uma nova LF-Tag expressão no. AWS Glue Data Catalog A expressão pode ser usada para conceder permissões aos recursos do Data Catalog, como bancos de dados, tabelas, visualizações ou colunas, com base nas tags associadas. Neste exemplo, a expressão coincidirá com recursos que têm a chave `Department` com os valores `Sales` ou `Marketing` e a chave `Location` com o valor `US`.

Como criador da expressão de tag, o diretor obtém `Alter` permissão sobre essa LF-Tag expressão e pode atualizar ou remover a expressão. O principal criador da LF-Tag expressão também pode conceder `Alter` permissão a outro diretor para atualizar e remover essa expressão.

Atualizando LF-Tag expressões

Somente administradores do data lake, o criador da LF-Tag expressão e os diretores que têm `Alter` ou têm `Super` permissão na LF-Tag expressão podem atualizar uma LF-Tag expressão. Além da permissão `Alter`, você também precisa da `lakeformation:UpdateLFTagExpression` do IAM e da `Grant with LF-Tag` em todos os valores de chave subjacentes no novo corpo da expressão para atualizar as expressões.

Você atualiza uma LF-Tag expressão atualizando a descrição, o corpo da expressão e as permissões concedidas na expressão. Você não pode alterar o nome da LF-Tag expressão. Para alterar o nome, exclua a LF-Tag expressão e adicione uma com os parâmetros necessários.

Você pode atualizar uma LF-Tag expressão usando o AWS Lake Formation console, a API ou o AWS Command Line Interface (AWS CLI).

Console

Para atualizar uma LF-Tag expressão

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

   Faça login como administrador do data lake, LF-Tag criador ou diretor com `Alter` permissão no LF-Tag.
2. No painel de navegação, em permissões, escolha LF-Tags e permissões.

3. Escolha a guia de LF-Tag expressões.
4. Na seção de LF-Tag expressões, selecione uma LF-Tag expressão e escolha Editar.
5. Na caixa de diálogo Editar LF-Tag expressão, atualize a descrição e o corpo da expressão adicionando ou removendo chaves e valores.

   Para adicionar vários valores, no campo Valores, escolha os valores no menu suspenso.
6. Escolha Salvar.

AWS CLI

O comando update-lf-tag-expression no Lake Formation permite que você atualize uma expressão existente. LF-Tag

```
aws lakeformation update-lf-tag-expression \
-- name expression_name\
-- description new_description \
-- catalog-id catalog_id \
-- expression '{"Expression": [{"TagKey": "tag_key", "TagValues": ["tag_value1",
 "tag_value2", ...]}]}'
```

Veja o que significam os parâmetros no comando fornecido:

- name: o nome da expressão de tag nomeada existente que você deseja atualizar.
- description: uma nova descrição para a expressão.

  catalog-id: o ID do Data Catalog em que a expressão de tag nomeada reside.
- expression: a nova string de expressão de tag com a qual você deseja atualizar a expressão.

Excluindo expressões LF-Tag

Você pode excluir LF-Tag expressões que não estão mais em uso. Se você concedeu permissões aos diretores nos recursos do Catálogo de Dados usando a LF-Tag expressão, eles não terão mais as permissões.

Somente administradores do data lake, o criador da LF-Tag expressão ou um diretor com `Drop` permissão na LF-Tag expressão podem excluir uma LF-Tag expressão. Além da `Drop` permissão,

o diretor também precisa da permissão `lakeformation:DeleteLFTagExpression` do IAM para excluir uma LF-Tag expressão.

Você pode excluir uma LF-Tag expressão usando o AWS Lake Formation console, a API ou o AWS Command Line Interface (AWS CLI).

Console

Para excluir uma LF-Tag expressão (console)

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

   Faça login como administrador do data lake, criador da LF-Tag expressão ou principal que tenha permissões para excluir a expressão.
2. No painel de navegação, em Permissões, escolha LF-Tags e permissões.
3. Escolha a guia de LF-Tag expressão.
4. Na seção de LF-Tag expressões, selecione uma LF-Tag expressão e escolha Excluir.
5. Na LF-Tag expressão Delete? caixa de diálogo, para confirmar a exclusão, insira o nome da LF-Tag expressão no campo designado e escolha Excluir.

AWS CLI

Para excluir um LF-Tag (AWS CLI)

- Insira um comando `delete-lf-tag-expression`. Forneça o nome da expressão e o ID do catálogo a serem excluídos.

  Example

  O exemplo a seguir exclui a LF-Tag expressão com o nome `my-tag-expression` do Catálogo de Dados com ID`123456789012`. O `catalog-id` parâmetro é opcional se você estiver usando a mesma conta da sua AWS CLI configuração. Depois de excluir uma LF-Tag expressão, o Lake Formation limpa os registros de permissão associados a essa expressão. Isso inclui registros de permissão individuais e registros de permissão agregados que contêm a expressão excluída.

  ```
  aws lakeformation delete-lf-tag-expression \
  --name "my-tag-expression" \
  ```

```
--catalog-id "123456789012"
```

Listando LF-Tag expressões

Você pode listar as LF-Tag expressões nas quais você tem as permissões Descrever. Administradores de data lake, criadores de LF-Tag expressões e Read-only administradores podem ver implicitamente todas as expressões de tag em suas contas.

Você pode listar LF-Tag expressões usando o AWS Lake Formation console, a API ou o AWS Command Line Interface (AWS CLI).

Console

Para listar LF-Tag expressões (console)

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

   Faça login como criador da LF-Tag expressão, como administrador do data lake ou como principal que recebeu permissões em LF-Tag expressões e que tem a permissão do `lakeformation:ListLFTagExpressions` IAM.
2. No painel de navegação, em Permissões LF-Tags e permissões.
3. Escolha a guia de LF-Tag expressões para ver as expressões. Esta seção mostra as informações sobre as LF-Tag expressões existentes, incluindo o nome da expressão, a própria expressão com links para as tags incluídas e as opções para criar, editar ou excluir expressões.

AWS CLI

Para listar LF-Tags (AWS CLI)

- Para listar LF-Tag expressões usando o AWS CLI, você pode usar o comando list-lf-tag-expressions. A sintaxe da solicitação é:

  ```
  aws lakeformation list-lf-tag-expressions \
  -- catalog-id "123456789012" \
  -- max-items "100" \
  -- next-token "next-token"
  ```

  Em que:

- `catalog-id`é o ID da AWS conta do Catálogo de Dados para o qual você deseja listar expressões de tag.
- `max-items` especifica o número máximo de expressões de tag a serem exibidas. Se esse parâmetro não é usado, o valor padrão é 100.
- `next-token` é um token de continuação caso os resultados tenham sido truncados em uma solicitação anterior.

A resposta incluirá uma lista de LF-Tag expressões e um próximo token, se aplicável.

Gerenciando permissões de LF-Tag valor

Você pode conceder as `Drop` `Alter` permissões LF-Tags aos diretores para gerenciar expressões de LF-Tag valor. Você também pode conceder `Describe` `Grant with LF-Tag expressions` permissões e permissões LF-Tags aos diretores para visualizá-los LF-Tags e atribuí-los aos recursos do Catálogo de Dados (bancos de dados, tabelas e colunas). `Associate` Quando LF-Tags são atribuídos aos recursos do Catálogo de Dados, você pode usar o método de controle de acesso (LF-TBAC) baseado em tags do Lake Formation para proteger esses recursos. Para obter mais informações, consulte Controle de acesso baseado em tags do Lake Formation.

Você pode conceder essas permissões com a opção de concessão para que outras entidades principais possam concedê-las. As permissões `Grant with LF-Tag expressions`, `Describe` e `Associate` são explicadas em Adicionar LF-Tag criadores.

Você pode conceder as `Associate` permissões `Describe` e em uma LF-Tag AWS conta externa. Um administrador de data lake nessa conta pode então conceder essas permissões a outras entidades principais na conta. Os diretores aos quais o administrador do data lake na conta externa concede a `Associate` permissão podem então atribuir LF-Tags aos recursos do Catálogo de Dados que você compartilhou com a conta deles.

Ao conceder para uma conta externa, você deve incluir a opção de concessão.

Você pode conceder permissões LF-Tags usando o console do Lake Formation, a API ou o AWS Command Line Interface (AWS CLI).

Tópicos

- Listando LF-Tag permissões usando o console
- Conceder LF-Tag permissões usando o console

- Gerenciando LF-Tag permissões usando o AWS CLI

Para obter mais informações, consulte Gerenciamento do controle LF-Tags de acesso aos metadados e Controle de acesso baseado em tags do Lake Formation.

Listando LF-Tag permissões usando o console

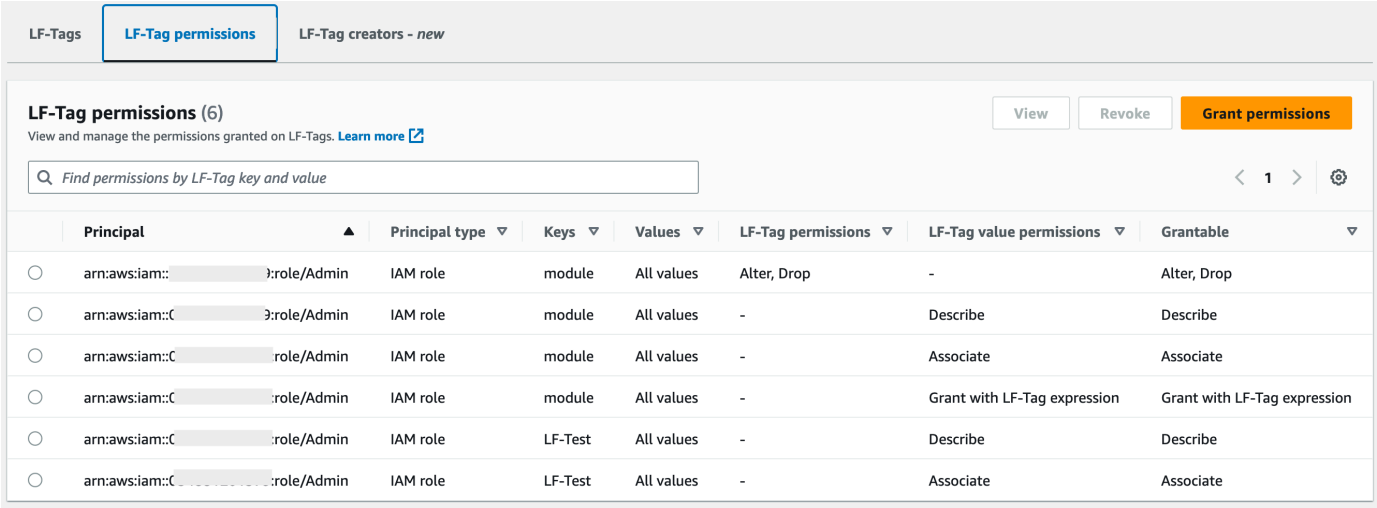
Você pode usar o console do Lake Formation para ver as permissões concedidas no LF-Tags. Você precisa ser um LF-Tag criador, administrador de um data lake ou ter a `Associate` permissão `Describe` ou a permissão de um LF-Tag para vê-lo.

Para listar LF-Tag permissões (console)

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

   Faça login como LF-Tag criador, administrador do data lake ou como usuário ao qual as `Describe` permissões `Drop` `AlterAssociate`,, ou LF-Tags foram concedidas.
2. No painel de navegação, em Permissões, escolha LF-Tags e permissões e escolha a seção de LF-Tag permissões.

   A seção de LF-Tag permissões mostra uma tabela que contém chaves principais, de tag, valores e permissões.



Conceder LF-Tag permissões usando o console

As etapas a seguir explicam como conceder permissões LF-Tags usando a página Conceder LF-Tag permissões no console do Lake Formation. A página está dividida nas seguintes seções:

- Tipos de permissão: o tipo de permissão a ser concedida.
- Entidades principais: os usuários ou perfis do IAM, ou os usuários ou perfis do SAML aos quais conceder permissões.
- LF-Tag permissões de pares de chave-valor — Os pares de LF-Tag valores-chave aos quais conceder permissões.
- LF-Tag permissões — A LF-Tags para conceder permissões.
- LF-Tag permissões de expressão — A LF-Tags para conceder permissões.
- Permissões – As permissões a serem concedidas.

Abra as Conceder LF-Tag permissões page

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

   Faça login como LF-Tag criador, administrador do data lake ou como usuário. LF-Tag Permissões de usuário ou par LF-Tag de valores-chave LF-Tags foram concedidas com a `Grant` opção.
2. No painel de navegação, escolha LF-Tags e permissões, escolha a seção de LF-Tagpermissões.
3. Escolha Conceder permissões.

Especifique o tipo de permissão

Na seção Tipo de permissões, escolha um tipo de permissão.

LF-Tag permissões

Escolha as LF-Tag permissões para permitir que os diretores atualizem ou LF-Tag LF-Tags excluam os valores.

LF-Tag permissões de pares de valores-chave

Escolha as permissões do par LF-Tag chave-valor para permitir que os diretores atribuam recursos, LF-Tags exibam LF-Tags e valores do Catálogo de Dados e concedam permissões LF-Tags baseadas nos recursos do Catálogo de Dados aos diretores.

As opções disponíveis nas seções a seguir dependem do Tipo de permissões.

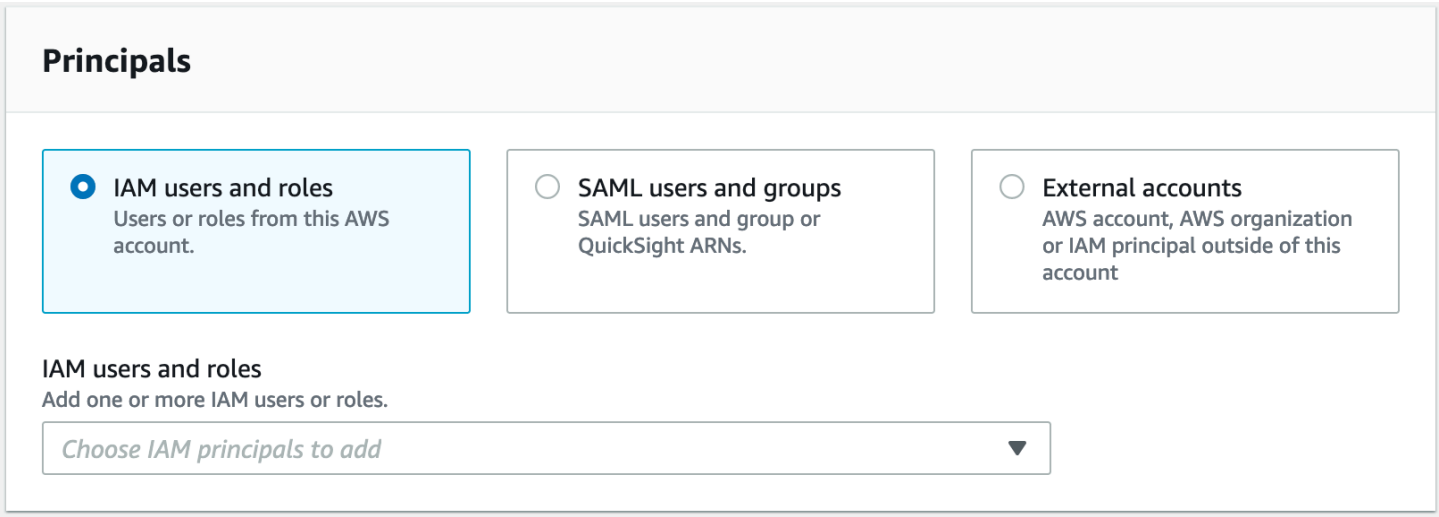
LF-Tag permissões de expressão

Escolha as permissões de LF-Tag expressão para permitir que os principais atualizem ou excluam expressões.

Especifique as entidades principais

> **Note**
>
> Você não pode conceder LF-Tag permissões (`AltereDrop`) a contas externas ou diretores em outra conta.

Na seção Entidades principais, escolha um tipo de entidade principal e especifique as entidades principais às quais conceder permissões.



Usuários e perfis do IAM

Escolha um ou mais usuários ou perfis na lista de usuários e perfis do IAM.

Usuários e grupos SAML

Para usuários e grupos SAML e Quick, insira um ou mais Amazon Resource Names (ARNs) para usuários ou grupos federados por meio de SAML ou ARNs para usuários ou grupos Quick. Pressione Enter após cada ARN.

Para obter informações sobre como construir os ARNs, consulte Concessão e revogação da Lake Formation AWS CLI Comandos da do.

> **Note**
>
> A integração do Lake Formation com o Quick é compatível somente com o Quick Enterprise Edition.

Contas externas

Em AWS conta, insira uma ou mais IDs de AWS conta válidas. Pressione Enter após cada ID.

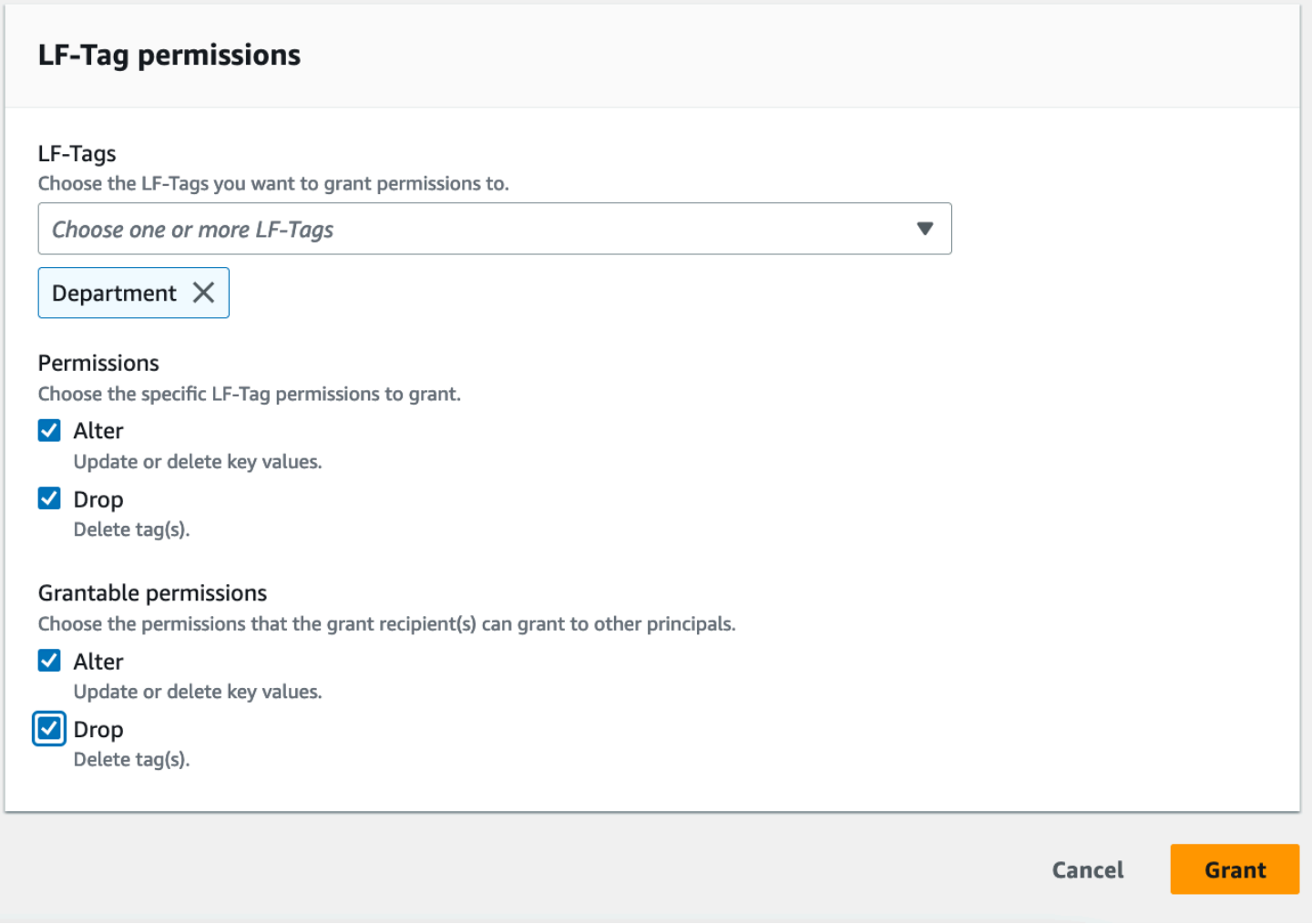
O ID da organização consiste em “o-” seguido por 10 a 32 letras minúsculas ou dígitos.

Um ID de unidade organizacional começa com “ou-” seguido de 4 a 32 letras minúsculas ou dígitos (o ID da raiz que contém a OU). Essa sequência é seguida por um segundo travessão "-" e 8 a 32 letras minúsculas ou dígitos adicionais.

Para a entidade principal do IAM, insira o ARN para o usuário ou o perfil do IAM.

Especifique o LF-Tags

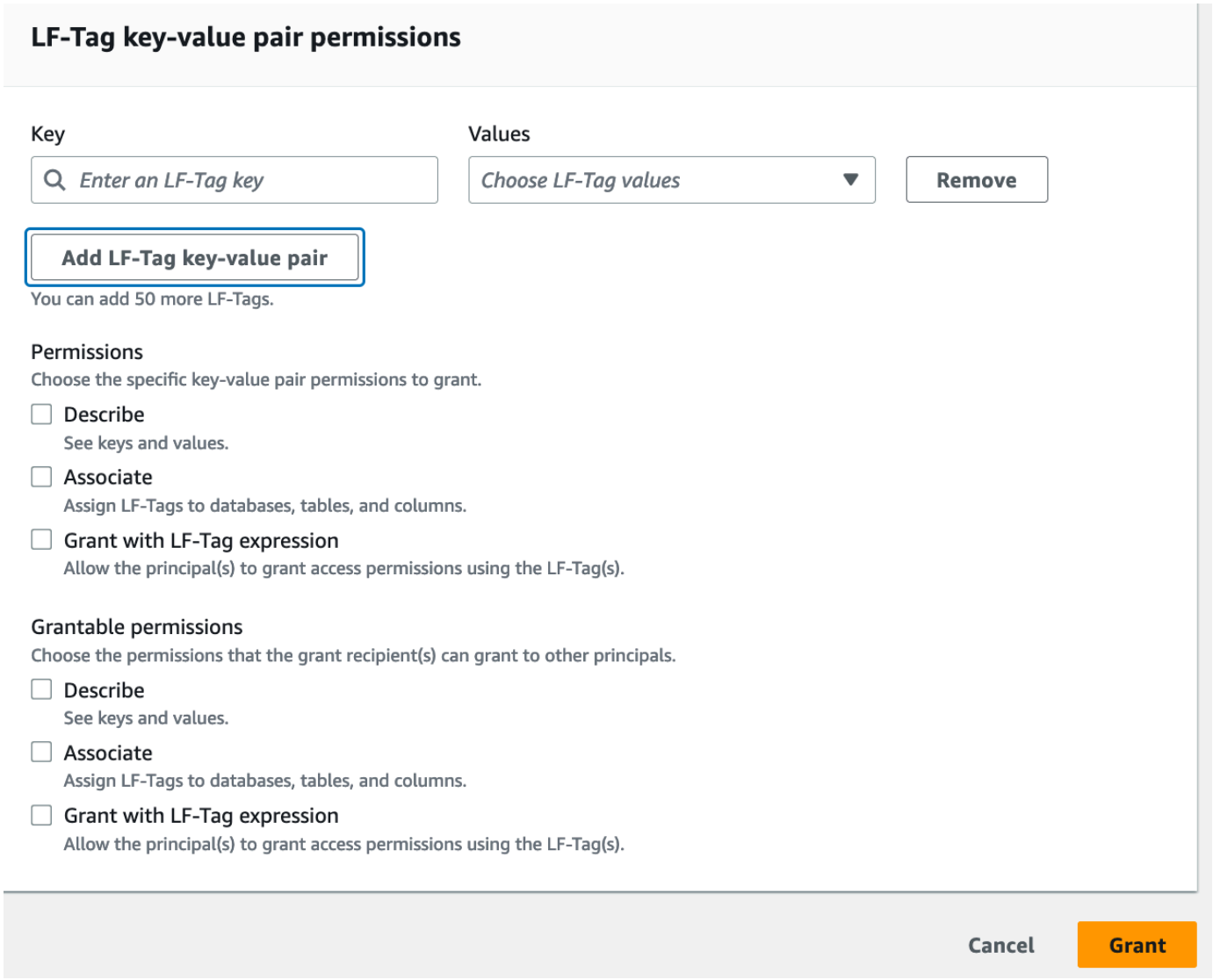
Para conceder permissões LF-Tags, na seção de LF-Tag permissões, especifique a opção LF-Tags para conceder permissões.



- Escolha um ou mais LF-Tag usando o menu suspenso.

Especifique os pares de LF-Tag valores-chave

1. Para conceder permissões em pares de LF-Tag chave-valor, (você precisa primeiro escolher as permissões do par LF-Tag chave-valor como o tipo de permissão), escolha Adicionar par LF-Tag chave-valor para revelar a primeira linha de campos para especificar a chave e os valores. LF-Tag



2. Posicione o cursor no campo Chave, opcionalmente comece a digitar para restringir a lista de seleção e selecione uma LF-Tag tecla.
3. Na lista Valores, selecione um ou mais valores e pressione Guia ou clique ou toque fora do campo para salvar os valores selecionados.

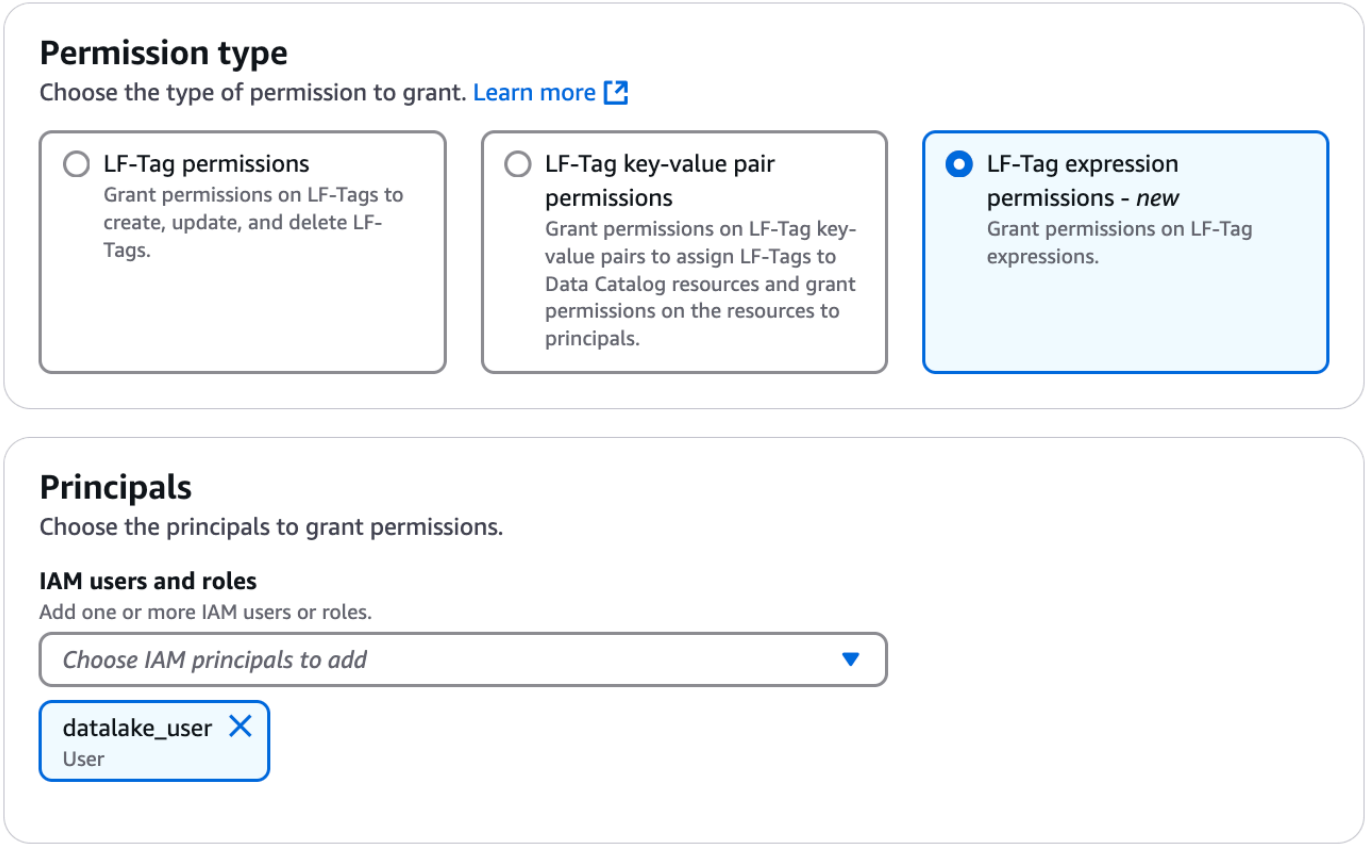
> **Note**
>
> Se uma das linhas na lista Valores estiver em foco, pressionar Enter marcará ou desmarcará a caixa de seleção.

Os valores selecionados aparecem como blocos abaixo da lista de Valores. Escolha o ✖ para remover um valor. Escolha Remover para remover o todo LF-Tag.

4. Para adicionar outro LF-Tag, escolha Adicionar LF-Tag novamente e repita as duas etapas anteriores.

Especifique as LF-Tag expressões

1. Para conceder permissões em LF-Tag expressões, (você precisa primeiro escolher escolher permissões de LF-Tag expressão como o tipo de permissão).



2. Escolha uma LF-Tag expressão.
3. As expressões selecionadas aparecem como blocos abaixo da lista de LF-Tag expressões. Clique no ✖ para remover uma expressão.
4. Para adicionar outra LF-Tag expressão, escolha outra expressão.

Especifique as permissões

Esta seção mostra as LF-Tag permissões ou as permissões de LF-Tag valor com base no tipo de permissão que você escolheu na etapa anterior.

Dependendo do tipo de permissão que você escolheu conceder, selecione LF-Tag as permissões ou as permissões do par LF-Tag de valores-chave e as permissões concedidas.

1. Em LF-Tag permissões, selecione as permissões a serem concedidas.

   Conceder Drop e Alter concede implicitamente Describe.

   Você precisa conceder as permissões Alterar e Eliminar em todos os valores da tag.
2. Em permissões LT-Tag de valor-chave, selecione as permissões a serem concedidas.

   Conceder Associate implicitamente concede Describe. Escolha Conceder com LF-Tag expressão para permitir que o destinatário da concessão conceda ou revogue permissões de acesso aos recursos do Catálogo de Dados usando o LF-TBAC método.
3. Em permissões de LF-Tag expressão, selecione as permissões a serem concedidas.

   Conceder Drop e Alter concede implicitamente Describe.

   Ao conceder a permissão Super, todas as permissões disponíveis são concedidas.
4. (Opcional) Em Permissões concedidas, selecione as permissões que o beneficiário do subsídio pode conceder a outros diretores em sua conta. AWS
5. Selecione Conceder.

Gerenciando LF-Tag permissões usando o AWS CLI

Você pode conceder, revogar e listar permissões LF-Tags usando o AWS Command Line Interface (AWS CLI).

Para listar LF-Tag permissões (AWS CLI)

- Insira um comando `list-permissions`. Você deve ser o LF-Tag criador, administrador do data lake ou ter a `Grant with LF-Tag permissions` permissão `Drop` `Alter` `DescribeAssociate`,,,, LF-Tag para vê-lo.

  O comando a seguir solicita tudo o LF-Tags que você tem permissões.

  ```
  aws lakeformation list-permissions --resource-type LF_TAG
  ```

  Veja a seguir um exemplo de saída para um administrador de data lake, que vê tudo LF-Tags concedido a todos os diretores. Non-administrative os usuários só veem LF-Tags como

concedido a eles. LF-Tag as permissões concedidas por uma conta externa aparecem em uma página de resultados separada. Para vê-los, repita o comando e forneça ao argumento `--next-token` o token retornado da execução anterior do comando.

```
{
    "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"
}
```

Você pode listar todas as concessões para uma LF-Tag chave específica. O comando a seguir retorna todas as permissões concedidas no LF-Tag `module`.

```
aws lakeformation list-permissions --resource-type LF_TAG --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

Você também pode listar LF-Tag valores concedidos a um principal específico para um determinado LF-Tag. Ao fornecer o argumento `--principal`, você deve fornecer o argumento `--resource`. Portanto, o comando só pode solicitar efetivamente os valores concedidos a um principal específico para uma LF-Tag chave específica. O comando a seguir mostra como fazer isso para o principal `datalake_user1` e a LF-Tag chave`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":["*"]}}'
```

Esta é uma saída de exemplo.

```
{
    "PrincipalResourcePermissions": [
        {
            "Principal": {
                "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"
            },
            "Resource": {
                "LFTag": {
                    "CatalogId": "111122223333",
                    "TagKey": "module",
                    "TagValues": [
```

```
                        "Orders",
                        "Sales"
                    ]
                }
            },
            "Permissions": [
                "ASSOCIATE"
            ],
            "PermissionsWithGrantOption": []
        }
    ]
}
```

Para conceder permissões em LF-Tags (AWS CLI)

1. Digite um comando semelhante ao seguinte: Este exemplo concede `datalake_user1` ao usuário a `Associate` permissão LF-Tag com a chave`module`. Ele concede permissões para visualizar e atribuir todos os valores dessa chave, conforme indicado pelo asterisco (*).

   ```
   aws lakeformation grant-permissions --principal
    DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
   datalake_user1 --permissions "ASSOCIATE" --resource '{ "LFTag":
    {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
   ```

   Conceder a permissão `Associate` implicitamente concede a permissão `Describe`.

   O próximo exemplo concede `Associate` à AWS conta externa 1234-5678-9012 com a chave, LF-Tag com a opção de concessão. `module` Ele concede permissões para visualizar e atribuir somente os valores `sales` e `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. Conceder a permissão `GrantWithLFTagExpression` implicitamente concede a permissão `Describe`.

O próximo exemplo `GrantWithLFTagExpression` concede a um usuário LF-Tag com a chave`module`, com a opção de concessão. Ele concede permissões para visualizar e conceder permissões nos recursos do catálogo de dados usando somente os valores `sales` e `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. O próximo exemplo concede `Drop` permissões a um usuário LF-Tag com a chave`module`, com a opção de concessão. Ele concede permissões para excluir LF-Tag o. Para excluir um LF-Tag, você precisa de permissões em todos os valores dessa chave.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333 --permissions "DROP"
 --permissions-with-grant-option "DROP" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

4. O próximo exemplo concede `Alter` permissões ao usuário LF-Tag com a chave`module`, com a opção de concessão. Ele concede permissões para excluir LF-Tag o. Para atualizar um LF-Tag, você precisa de permissões em todos os valores dessa chave.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333 --permissions "ALTER"
 --permissions-with-grant-option "ALTER" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

Para revogar permissões em LF-Tags (AWS CLI)

- Digite um comando semelhante ao seguinte: Este exemplo revoga a `Associate` permissão LF-Tag com a chave `module` do usuário`datalake_user1`.

```
aws lakeformation revoke-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --permissions "ASSOCIATE" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

Concedendo permissões de data lake usando o método LF-TBAC

Você pode conceder permissões de `DESCRIBE` e `ASSOCIATE` Lake Formation LF-Tags aos diretores para que eles possam visualizá-los LF-Tags e atribuí-los aos recursos do Catálogo de Dados (bancos de dados, tabelas, visualizações e colunas). Quando LF-Tags são atribuídos aos recursos do Catálogo de Dados, você pode usar o método de controle de acesso (LF-TBAC) baseado em tags do Lake Formation para proteger esses recursos. Para obter mais informações, consulte Controle de acesso baseado em tags do Lake Formation.

No início, somente o administrador do data lake pode conceder essas permissões. Se o administrador do data lake conceder essas permissões com a opção de concessão, outras entidades principais poderão concedê-las. As permissões `DESCRIBE` e `ASSOCIATE` são explicadas em Considerações e práticas recomendadas de controle de acesso com base em tags do Lake Formation.

Você pode conceder as `ASSOCIATE` permissões `DESCRIBE` e em uma LF-Tag AWS conta externa. Um administrador de data lake nessa conta pode então conceder essas permissões a outras entidades principais na conta. Os diretores aos quais o administrador do data lake na conta externa concede a `ASSOCIATE` permissão podem então atribuir LF-Tags aos recursos do Catálogo de Dados que você compartilhou com a conta deles.

Ao conceder para uma conta externa, você deve incluir a opção de concessão.

Você pode conceder permissões LF-Tags usando o AWS Lake Formation console, a API ou o AWS Command Line Interface (AWS CLI).

> **Note**
>
> As etapas a seguir não são necessárias para catálogos de Tabelas do S3. Você pode usar LF-Tags para conceder permissões em catálogos existentes do S3 Tables sem excluí-los e recriá-los.

Habilitando o LF-Tags suporte para catálogos federados existentes que usam as permissões do Lake Formation

Siga estas etapas, se você tiver catálogos federados existentes que estejam usando permissões do Lake Formation, como o Amazon Redshift Amazon DynamoDB ou catálogos que foram criados LF-Tags antes que o suporte estivesse disponível para catálogos federados.

1. Exclua o catálogo existente: chame a operação `deleteCatalog` da API para remover o catálogo federado existente que usa as permissões do Lake Formation.
2. Crie um novo catálogo federado — Crie um novo catálogo e direcione o novo catálogo para o existente namespace/datashare.

   Use um novo nome para o catálogo — Esse processo atualiza seus catálogos federados preexistentes para oferecer suporte à funcionalidade. LF-Tag Se você quiser usar o mesmo nome de catálogo, entre em contato com a equipe de AWS suporte para obter ajuda.

Tópicos

- Conceder permissões do catálogo de dados

Consulte também

- Gerenciando permissões de LF-Tag valor
- Gerenciamento do controle LF-Tags de acesso aos metadados
- Controle de acesso baseado em tags do Lake Formation

Conceder permissões do catálogo de dados

Use o console do Lake Formation ou AWS CLI conceda permissões do Lake Formation em bancos de dados, tabelas, visualizações e colunas do Catálogo de Dados usando o método de controle de acesso (LF-TBAC) baseado em tags do Lake Formation.

Console

As etapas a seguir explicam como conceder permissões usando o método de controle de acesso (LF-TBAC) baseado em tags do Lake Formation e a página Conceder permissões do data lake no console do Lake Formation. A página está dividida nas seguintes seções:

- Diretores — Os usuários, as funções e Contas da AWS para os quais conceder permissões.
- LF-Tags ou recursos do catálogo — Os bancos de dados, tabelas ou links de recursos nos quais conceder permissões.
- Permissões – As permissões do Lake Formation devem ser concedidas.

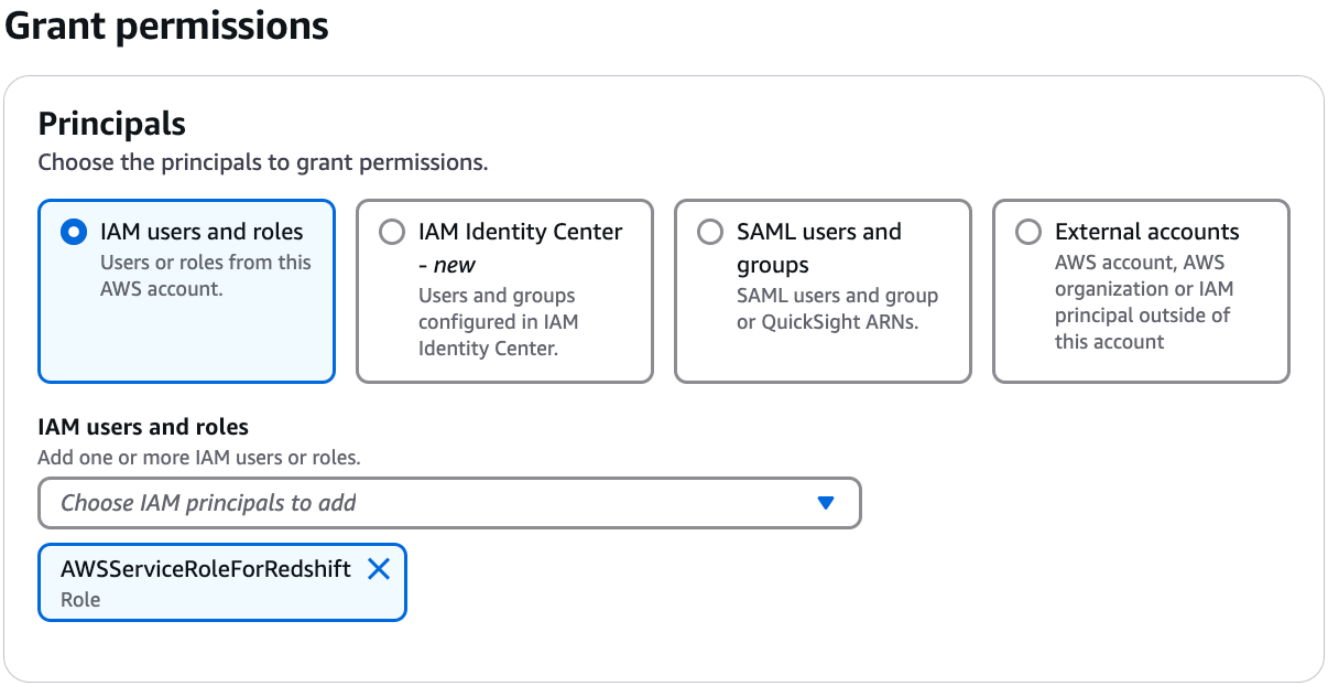
1. Abra a página Conceder permissões de data lake.

   Abra o AWS Lake Formation console em https://console.aws.amazon.com/lakeformation/e faça login como administrador do data lake ou como usuário que recebeu permissões do Lake Formation nos recursos do Catálogo de Dados por meio LF-TBAC da opção de concessão.

   No painel de navegação, em Permissões, escolha Permissões do data lake. Em seguida, escolha Conceder.

2. Especifique as entidades principais.

   Na seção Entidades principais, escolha um tipo de principal e, em seguida, especifique às entidades principais aos quais conceder permissões.



   Usuários e perfis do IAM

   Escolha um ou mais usuários ou perfis na lista de usuários e perfis do IAM.

   Centro de Identidade do IAM

   Selecione um ou mais usuários na lista Usuários e grupos.

Usuários e grupos SAML

Para usuários e grupos SAML e Quick, insira um ou mais Amazon Resource Names (ARNs) para usuários ou grupos federados por meio de SAML ou ARNs para usuários ou grupos Quick. Pressione Enter após cada ARN.

Para obter informações sobre como construir os ARNs, consulte Concessão e revogação da Lake Formation AWS CLI Comandos da do.

> **Note**
>
> A integração do Lake Formation com o Quick é compatível somente com o Quick Enterprise Edition.

Contas externas

Para Contas da AWS, AWS organização ou diretor do IAM, insira um ou mais Conta da AWS IDs válidos, IDs da organização, IDs da unidade organizacional ou ARN para o usuário ou a função do IAM. Pressione Enter após cada ID.

O ID da organização consiste em “o-” seguido por 10 a 32 letras minúsculas ou dígitos.

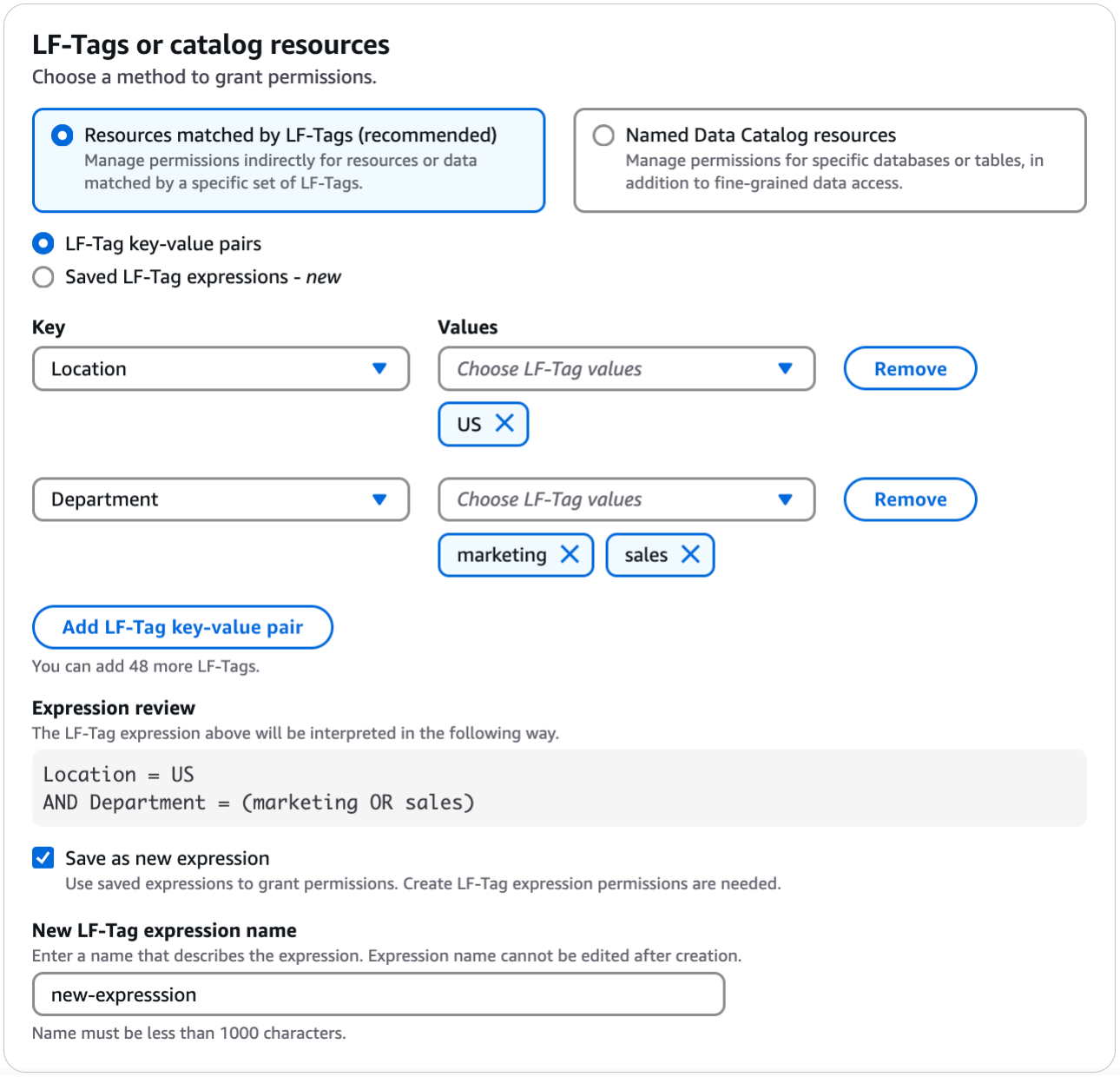
Um ID de unidade organizacional começa com “ou-” seguido de 4 a 32 letras minúsculas ou dígitos (o ID da raiz que contém a OU). Essa sequência é seguida por um segundo travessão "-" e 8 a 32 letras minúsculas ou dígitos adicionais.

3. Especifique LF-Tags o.

Certifique-se de que a LF-Tags opção Recursos correspondidos por seja escolhida. Escolha pares LF-Tag de valores-chave ou expressões salvas LF-Tag .

1. Se você escolher a opção de pares LF-Tag de valores-chave, escolha as chaves e os valores.

   Se você escolher mais de um valor, estará criando uma LF-Tag expressão com um `OR` operador. Isso significa que, se algum dos LF-Tag valores corresponder a um recurso LF-Tag atribuído a um catálogo de dados, você receberá permissões sobre o recurso.



2. (Opcional) Escolha Adicionar par LF-Tag de valores-chave novamente para especificar outro. LF-Tag

   Se você especificar mais de um LF-Tag, estará criando uma LF-Tag expressão com um AND operador. O principal recebe permissões em um recurso do Catálogo de Dados somente se o recurso tiver recebido uma correspondência LF-Tag para cada um LF-Tag na LF-Tag expressão.

3. Escolha a opção Salvar como nova expressão para reutilizar a expressão.

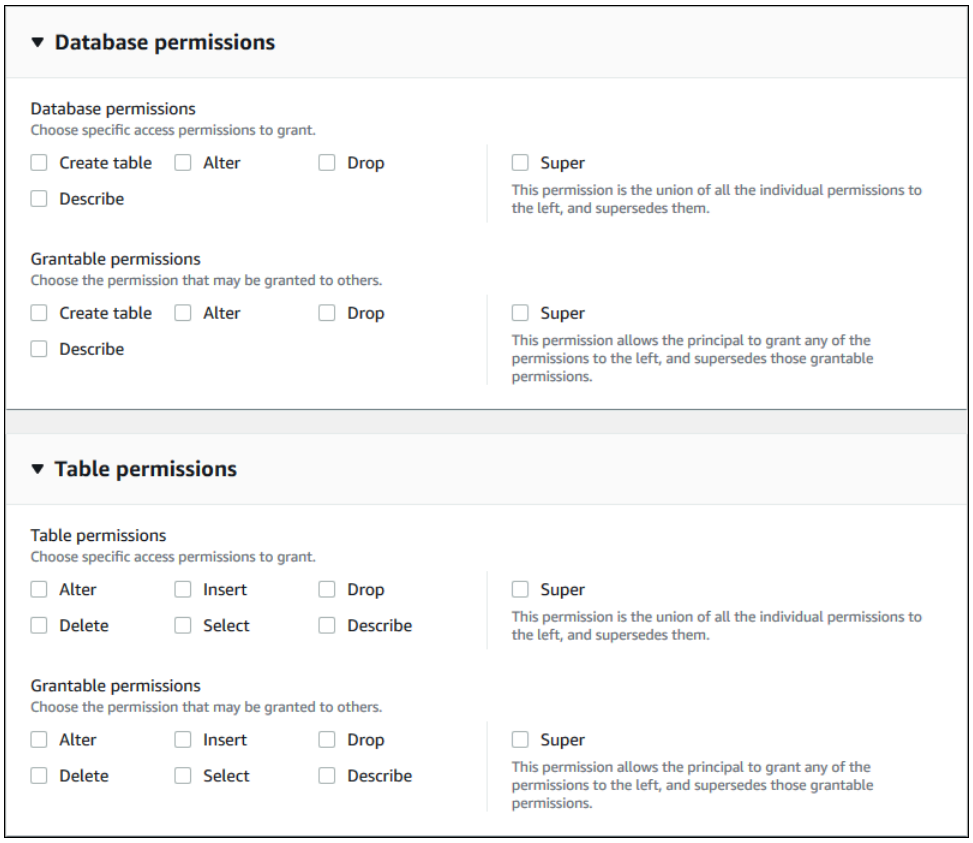
   Você precisa de `Create LF-Tag expression` para salvar as expressões.

Para obter mais informações sobre LF-Tag expressões, consulte[Gerenciando LF-Tag expressões para controle de acesso a metadados](#).

4. Especifique as permissões.

   Especificar as permissões a serem concedidas à entidade principal em recursos correspondentes do catálogo de dados. Recursos correspondentes são aqueles que foram atribuídos LF-Tags e correspondem a uma das LF-Tag expressões concedidas ao diretor.

   É possível especificar as permissões a serem concedidas em bancos de dados, tabelas e visualizações correspondentes.



   Em Permissões do banco de dados, selecione as permissões do banco de dados a serem concedidas à entidade principal nos bancos de dados correspondentes.

   Em Permissões de tabela, selecione as permissões de tabela ou visualização a serem concedidas à entidade principal nas tabelas e nas visualizações correspondentes.

   Também é possível selecionar as permissões `Select`, `Describe` e `Drop` nas Permissões de tabela a serem aplicadas às visualizações.

5. Selecione Conceder.

AWS CLI

Você pode usar o AWS Command Line Interface (AWS CLI) e o método de controle de acesso (LF-TBAC) baseado em tags do Lake Formation para conceder permissões do Lake Formation nos bancos de dados, tabelas e colunas do Catálogo de Dados.

Concedendo permissões de data lake usando o AWS CLI e o LF-TBAC método

- Use o comando `grant-permissions`.

  Example

  O exemplo a seguir concede a LF-Tag expressão `module=*` "" (todos os valores da LF-Tag chave`module`) ao usuário`datalake_user1`. Esse usuário terá a `CREATE_TABLE` permissão em todos os bancos de dados correspondentes — bancos de dados que foram atribuídos LF-Tag com a chave`module`, com qualquer valor.

```
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

  O próximo exemplo concede a LF-Tag expressão "`(level=director) AND (region=west OR region=south)`" ao usuário`datalake_user1`. Esse usuário terá as permissões `SELECT`, `ALTER` e `DROP` com a opção de concessão em tabelas correspondentes – tabelas que foram atribuídas tanto a `level=director` quanto a `region=west` ou `region=south`.

```
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

O próximo exemplo concede a LF-Tag expressão "`module=orders`" à AWS conta 1234-5678-9012. O administrador do data lake nessa conta pode então conceder a expressão "`module=orders`" às entidades principais em sua conta. Esses diretores terão então a `CREATE_TABLE` permissão para combinar bancos de dados pertencentes à conta 1111-2222-3333 e compartilhados com a conta 1234-5678-9012 usando o método de recurso nomeado ou o método. 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"]}]}}'
```

Attribute-based controle de acesso

No AWS Lake Formation, você pode conceder acesso a AWS Glue Data Catalog objetos como catálogos, bancos de dados, tabelas e filtros de dados usando atributos que são tags do IAM e tags de sessão associadas a entidades do IAM, como funções e usuários.

Para obter mais informações sobre o uso de tags de sessão, consulte assume-role no guia do AWS CLI usuário.

Attribute-based controle de acesso (ABAC) é uma estratégia de autorização que define permissões com base em atributos. AWS chama esses atributos de tags. É possível usar o ABAC para conceder acesso às entidades principais na mesma conta ou em outra conta nos recursos do Data Catalog. Qualquer entidade principal do IAM com chaves e valores correspondentes de tag do IAM ou de tag de sessão recebe acesso ao recurso. É necessário ter permissões para concessão nos recursos para fazer essas concessões.

O ABAC permite que você conceda acesso a vários usuários ao mesmo tempo. Quando novos usuários ingressam na organização, o acesso deles aos dados pode ser determinado automaticamente com base nos respectivos atributos, como cargo ou departamento, sem exigir que os administradores atribuam manualmente permissões ou perfis específicos. Ao usar atributos em vez de perfis, o ABAC oferece uma maneira mais simplificada e sustentável de gerenciar o acesso aos dados em diversos sistemas e ambientes, melhorando, em última análise, a governança e a conformidade dos dados.

Para acessar mais informações sobre como definir atributos, consulte Definir permissões com base em atributos com autorização ABAC.

Para obter informações sobre limitações, considerações e AWS regiões suportadas, consulteConsiderações sobre controle de acesso por atributo, limitações e regiões aceitas.

Tópicos

- Pré-requisitos para conceder permissões usando atributos
- Conceder permissões usando o controle de acesso por atributo

Pré-requisitos para conceder permissões usando atributos

Para conceder permissões usando o controle de acesso por atributo (ABAC), você deve atender aos seguintes pré-requisitos:

- Atualize as configurações do Data Catalog para habilitar as permissões do Lake Formation para objetos do Data Catalog. Para acessar mais informações, consulte a seção Alterar o modelo de permissão padrão ou usar o modo de acesso híbrido.
- Defina as configurações da versão entre contas como duas ou mais.
- Anexe atributos às entidades do IAM que exigem acesso.
- Somente um administrador do data lake ou um usuário do IAM com as permissões necessárias pode conceder acesso aos objetos do Data Catalog. Para acessar mais informações sobre as permissões necessárias, consulte Permissões do IAM.

Conceder permissões usando o controle de acesso por atributo

Este tópico descreve as etapas que você precisa seguir para conceder permissões de acesso por atributo em recursos do Data Catalog. Você pode usar o console do Lake Formation ou a Interface de Linha de AWS Comando (AWS CLI).

Concedendo permissões usando ABAC (Console de gerenciamento da AWS)

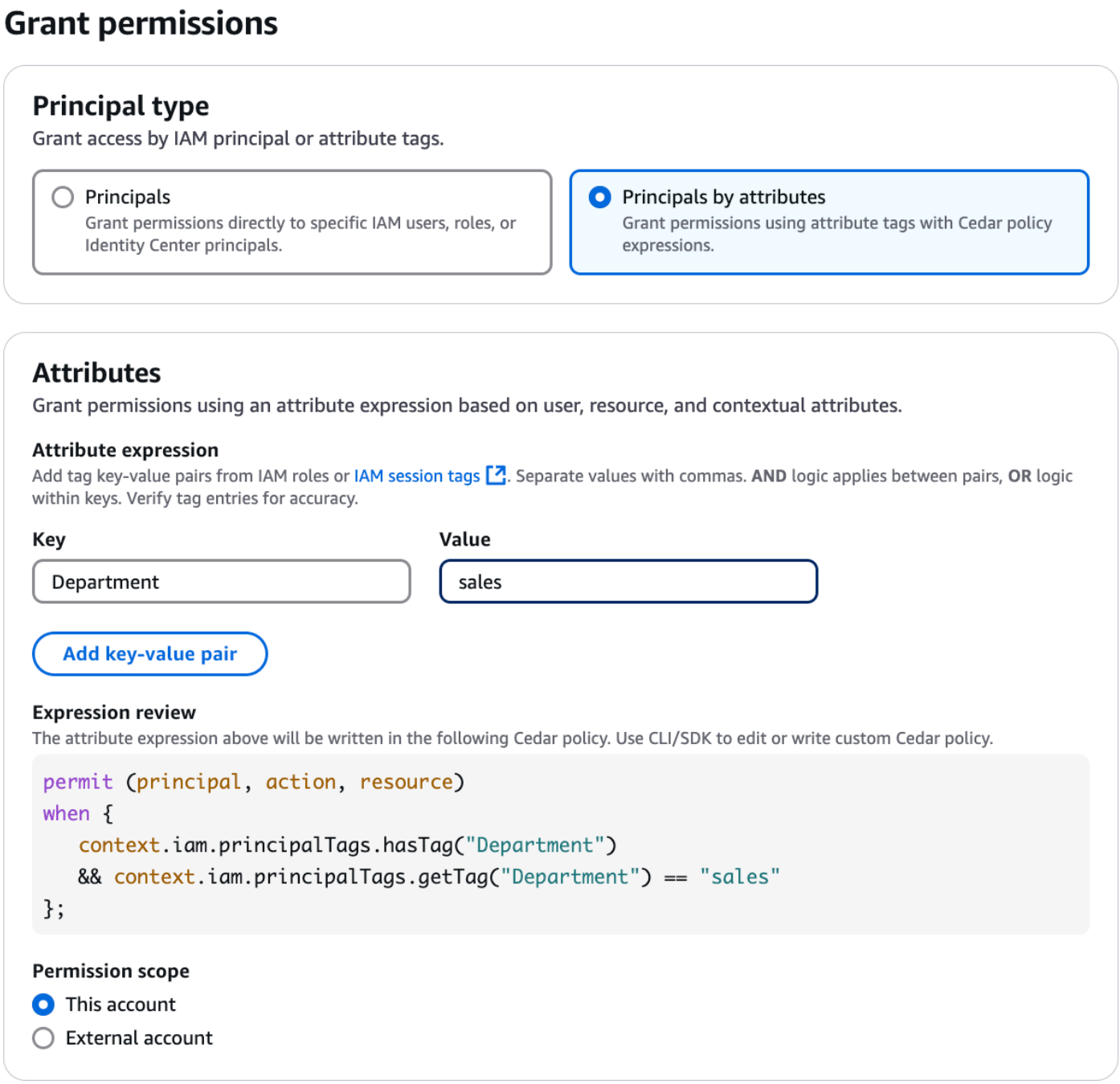
1. Abra o console do Lake Formation em e faça login como administrador do data lake, criador do recurso ou usuário do IAM com permissões concedidas no recurso. https://console.aws.amazon.com/lakeformation/
2. Execute um destes procedimentos:

- No painel de navegação, em Permissões, escolha Permissões do data lake. Em seguida, escolha Conceder.
- No painel de navegação, escolha Catálogos, em Data Catalog. Depois, escolha um objeto de catálogo (catálogos, bancos de dados, tabelas e filtros de dados) e, no menu Ações, em Permissões, escolha Conceder.

3. Na página Conceder permissões, escolha Entidades principais por atributo.
4. Especifique a chave e os valores do atributo. Se você escolher mais de um valor, estará criando uma expressão de atributo com um operador `OR`. Isso significa que, se algum dos valores da tag de atributo atribuídos a uma função ou usuário do IAM corresponder, eles role/user ganharão permissões de acesso ao recurso.

   Se você especificar mais de uma tag de atributo, estará criando uma expressão de atributo com um operador `AND`. O principal recebe permissões em um recurso do Catálogo de Dados somente se o IAM tiver role/user recebido uma tag correspondente para cada tag de atributo na expressão do atributo.

   Analise a expressão da política Cedar resultante mostrada no console.



5. Escolha o escopo da permissão. Se os beneficiários pertencerem a uma conta externa, escolha Conta externa e insira o ID da AWS conta.
6. Depois, escolha a conta do Data Catalog ou contas externas. Você deve ter as permissões para concessão correspondentes nos recursos para concluir com êxito as concessões de permissão.
7. Especifique quais ações você deseja permitir que as entidades principais (usuários ou perfis) que têm atributos correspondentes executem. O acesso é concedido a entidades do IAM às quais

foram atribuídas tags e valores que correspondem a pelo menos uma das expressões de atributo especificadas. Analise a expressão da política Cedar no console. Para acessar mais informações sobre Cedar, consulte O que é Cedar? | Referência GuideLink de linguagem da Cedar Policy.

8. Depois, escolha os recursos do Data Catalog para conceder acesso. Você pode definir essas permissões para vários recursos do Data Catalog, incluindo catálogos, bancos de dados, tabelas e filtros de dados.
9. Selecione Conceder.

   Essa abordagem permite controlar o acesso por atributo, garantindo que somente usuários ou perfis com as tags apropriadas possam realizar ações específicas nos recursos designados.

Concedendo permissões usando ABAC (AWS CLI)

O exemplo a seguir mostra uma expressão de atributo que deve ser atendida para receber todas as permissões disponíveis no recurso. Você também pode especificar permissões individuais, como `Select`, `Describe` ou `Drop`. A expressão usa a expressão de política Cedar. Para acessar mais informações sobre Cedar, consulte O que é Cedar? | Referência GuideLink de linguagem da Cedar Policy.

Essa condição confere se a entidade principal do IAM tem uma tag `department` e o valor da tag `department` é igual a `sales`.

```
aws lakeformation grant-permissions
--principal '{"DataLakePrincipalIdentifier": "111122223333:IAMPrincipals"}' \
--resource '{"Database": {"CatalogId": 111122223333, "Name": "abac-db"}}' \
--permissions ALL \
--condition '{"Expression": "context.iam.principalTags.hasTag(\"department\") \
   && context.iam.principalTags.getTag(\"department\") == \"sales\""}'
```

Exemplo de cenário de permissões

O cenário a seguir ajuda a demonstrar como você pode configurar permissões para proteger o acesso aos dados no AWS Lake Formation.

Shirley é administradora de dados. Ela quer configurar um data lake para sua empresa, AnyCompany. Atualmente, todos os dados são armazenados no Amazon S3. John é gerente de marketing e precisa ter acesso por escrito às informações de compra do cliente (contidas em `s3://`

`customerPurchases`). Um analista de marketing, Diego, se junta a John neste verão. John precisa da capacidade de conceder acesso a Diego para realizar consultas nos dados sem envolver Shirley.

Mateo, do setor financeiro, precisa acessar dados contábeis de consulta (por exemplo, `s3://transactions`). Ele quer consultar os dados das transações em tabelas em um banco de dados (`Finance_DB`) que a equipe financeira usa. Seu gerente, Arnav, pode dar a ele acesso ao `Finance_DB`. Embora ele não deva ser capaz de modificar os dados contábeis, ele precisa converter os dados em um formato (esquema) adequado para previsões. Esses dados serão armazenados em um bucket separado (`s3://financeForecasts`) que ele pode modificar.

Para resumir:

- Shirley é a administradora do data lake.
- John exige permissão `CREATE_DATABASE` e `CREATE_TABLE` para criar novos bancos de dados e tabelas no catálogo de dados.
- John também exige `SELECT`, `INSERT` e `DELETE` permissões nas tabelas que ele cria.
- Diego exige permissão `SELECT` na tabela para executar consultas.

Os funcionários da AnyCompany realizam as seguintes ações para configurar as permissões. As operações de API mostradas nesse cenário mostram uma sintaxe simplificada para maior clareza.

1. Shirley registra o caminho do Amazon S3 contendo informações de compra do cliente no Lake Formation.

   ```
   RegisterResource(ResourcePath("s3://customerPurchases"), false, Role_ARN )
   ```

2. Shirley concede a John o acesso ao caminho do Amazon S3 contendo informações de compra do cliente.

   ```
   GrantPermissions(John, S3Location("s3://customerPurchases"),
    [DATA_LOCATION_ACCESS]) )
   ```

3. Shirley concede permissão a John para criar bancos de dados.

   ```
   GrantPermissions(John, catalog, [CREATE_DATABASE])
   ```

4. John cria o banco de dados `John_DB`. John tem permissão `CREATE_TABLE` automática nesse banco de dados porque ele o criou.

```
CreateDatabase(John_DB)
```

5. John cria a tabela `John_Table` apontando para `s3://customerPurchases`. Como ele criou a tabela, ele tem todas as permissões nela e pode conceder permissões sobre ela.

```
CreateTable(John_DB, John_Table)
```

6. John permite que seu analista, Diego, tenha acesso à tabela `John_Table`.

```
GrantPermissions(Diego, John_Table, [SELECT])
```

7. John permite que seu analista, Diego, acesse o `s3://customerPurchases/London/`. Como Shirley já está registrada `s3://customerPurchases`, suas subpastas são registradas no Lake Formation.

```
GrantDataLakePrivileges( 123456789012/datalake, Diego, [DATA_LOCATION_ACCESS], [],
S3Location("s3://customerPurchases/London/") )
```

8. John permite que seu analista, Diego, crie tabelas no banco de dados `John_DB`.

```
GrantDataLakePrivileges( 123456789012/datalake, Diego, John_DB, [CREATE_TABLE],
[] )
```

9. Diego cria uma tabela `John_DB` em `s3://customerPurchases/London/` e obtém automaticamente as permissões `ALTER`, `DROP`, `SELECT`, `INSERT` e `DELETE`.

```
CreateTable( 123456789012/datalake, John_DB, Diego_Table )
```

Filtragem de dados e segurança por célula no Lake Formation

Ao conceder permissões do Lake Formation em uma tabela do catálogo de dados, você pode incluir especificações de filtragem de dados para restringir o acesso a determinados dados nos resultados da consulta e nos mecanismos integrados ao Lake Formation. O Lake Formation usa a filtragem de dados para obter segurança por coluna, segurança por linha e segurança por célula. Será possível definir e aplicar filtros de dados em colunas aninhadas se os dados de origem contiverem estruturas aninhadas.

Com os recursos de filtragem de dados do Lake Formation, você pode implementar os seguintes níveis de segurança de dados.

Column-level segurança

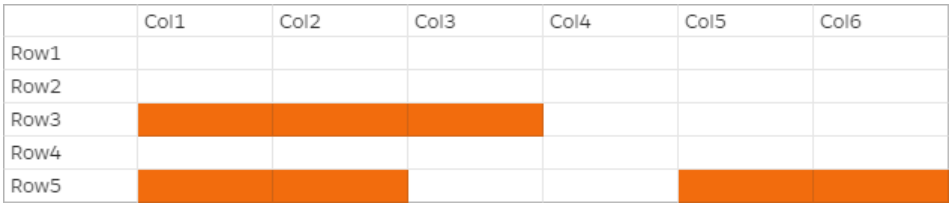
Conceder permissões em uma tabela do catálogo de dados com segurança em nível de coluna (filtragem de colunas) permite que os usuários visualizem somente colunas específicas e colunas aninhadas às quais eles têm acesso na tabela. Considere uma tabela do `persons` usada em vários aplicativos para uma grande empresa de comunicações multirregional. A concessão de permissões em tabelas do catálogo de dados com filtragem de colunas pode impedir que usuários que não trabalham no departamento de RH vejam informações de identificação pessoal (PII), como número de previdência social ou data de nascimento. Também é possível definir políticas de segurança e conceder acesso somente a subestruturas parciais de colunas aninhadas.

Row-level segurança

A concessão de permissões em uma tabela do catálogo de dados com segurança por linha (filtragem de linha) permite que os usuários visualizem somente linhas específicas de dados às quais eles têm acesso na tabela. A filtragem é baseada nos valores de uma ou mais colunas. É possível incluir estruturas de colunas aninhadas ao definir expressões de filtro de linha. Por exemplo, se diferentes escritórios regionais da empresa de comunicações tiverem seus próprios departamentos de RH, você poderá limitar os registros pessoais que os funcionários de RH podem ver a apenas registros de funcionários em sua região.

Cell-level segurança

Cell-level a segurança combina filtragem de linhas e filtragem de colunas para um modelo de permissões altamente flexível. Se você exibir as linhas e colunas de uma tabela como uma grade, usando a segurança por célula, poderá restringir o acesso a elementos individuais (células) da grade em qualquer lugar nas duas dimensões. Ou seja, você pode restringir o acesso a diferentes colunas, dependendo da linha. Veja o diagrama a seguir, no qual as colunas restritas são sombreadas.



Continuando com o exemplo da tabela de pessoas, você pode criar um filtro de dados no nível da célula que restringe o acesso à coluna de endereço da rua se a linha tiver a coluna do país definida

como “Reino Unido”, mas permite o acesso à coluna do endereço da rua se a linha tiver a coluna do país definida como “EUA”.

Os filtros se aplicam somente a operações de leitura. Portanto, você pode conceder somente a permissão `SELECT` do Lake Formation com filtros.

Cell-level segurança em colunas aninhadas

O Lake Formation permite definir e aplicar filtros de dados com segurança em nível de célula em colunas aninhadas. No entanto, os mecanismos analíticos integrados, como Amazon Athena, Amazon EMR e Amazon Redshift Spectrum, aceitam a execução de consultas em tabelas aninhadas gerenciadas pelo Lake Formation com segurança em nível de linha e coluna.

Para conhecer as limitações, consulte Limitações de filtragem de dados.

Tópicos

- Filtros de dados no Lake Formation
- Suporte PartiQL em expressões de filtro de linha
- Permissões necessárias para consultar tabelas com filtragem em nível de célula
- Como gerenciar filtros de dados

Filtros de dados no Lake Formation

É possível implementar segurança por coluna, por linha e por célula ao criar filtros de dados. Você seleciona um filtro de dados ao conceder permissão `SELECT` ao Lake Formation nas tabelas. Se a tabela contiver estruturas de colunas aninhadas, será possível definir um filtro de dados incluindo ou excluindo as colunas secundárias e definir expressões de filtro em nível de linha em atributos aninhados.

Cada filtro de dados pertence a uma tabela específica em seu catálogo de dados. Um filtro de dados inclui as seguintes informações:

- Nome do filtro
- Os IDs de catálogo da tabela associada ao filtro
- Nome da tabela
- Nome do banco de dados que contém a tabela

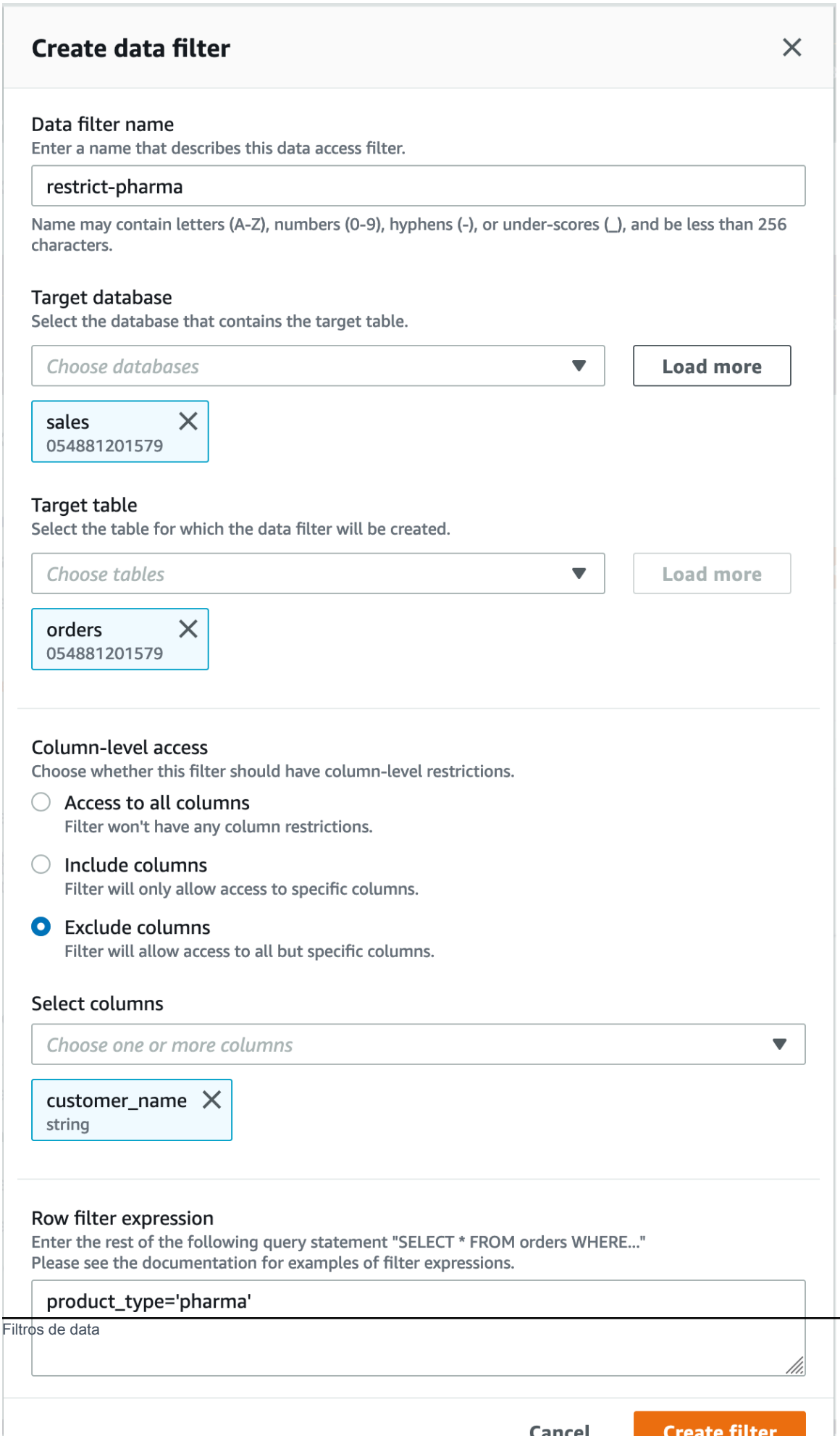
- Especificação de coluna: uma lista de colunas e colunas aninhadas (com tipos de dados `struct`) a serem incluídas ou excluídas dos resultados da consulta.
- Expressão de filtro de linha — uma expressão que especifica as linhas a serem incluídas nos resultados da consulta. Com algumas restrições, a expressão tem a sintaxe de uma cláusula `WHERE` na linguagem PartiQL. Para especificar todas as linhas, escolha Acesso a todas as linhas em Row-level Acesso no console ou Uso `AllRowsWildcard` em chamadas de API.

  Para obter mais informações sobre o que é suportado em expressões de filtro de linha, consulte Suporte PartiQL em expressões de filtro de linha.

O nível de filtragem obtido depende de como você preenche o filtro de dados.

- Ao especificar o curinga “todas as colunas” e fornecer uma expressão de filtro de linha, você está estabelecendo apenas a segurança por linha (filtragem de linhas).
- Ao incluir ou excluir colunas específicas e colunas aninhadas e especificar “todas as linhas” utilizando o curinga de todas as linhas, você está estabelecendo somente a segurança em nível de coluna (filtragem de colunas).
- Quando você inclui ou exclui colunas específicas e também fornece uma expressão de filtro de linha, está estabelecendo a segurança por célula (filtragem de células).

A captura de tela a seguir do console do Lake Formation mostra um filtro de dados que realiza a filtragem em nível de célula. Para consultas na tabela `orders`, ela restringe o acesso à coluna `customer_name` e os resultados da consulta retornam somente as linhas em que a coluna `product_type` contém “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
Row filter expression
Enter the rest of the following query statement "SELECT * FROM orders WHERE..."
Please see the documentation for examples of filter expressions.
product_type='pharma'
Cancel
Create filter

Observe o uso de aspas simples para incluir o literal da string de caracteres, `'pharma'`.

Você pode usar o console do Lake Formation para criar esse filtro de dados ou fornecer o seguinte objeto de solicitação para a operação `CreateDataCellsFilter` da API.

```
{
     "Name": "restrict-pharma",
     "DatabaseName": "sales",
     "TableName": "orders",
     "TableCatalogId": "111122223333",
     "RowFilter": {"FilterExpression": "product_type='pharma'"},
     "ColumnWildcard": {
         "ExcludedColumnNames": ["customer_name"]
     }
}
```

É possível criar tantos filtros de dados quantos necessários para uma tabela. Para fazer isso, você precisa de permissão SELECT com a opção de concessão em uma tabela. Por padrão, os administradores do Data Lake têm permissão para criar filtros de dados em todas as tabelas dessa conta. Normalmente, você usa apenas um subconjunto dos filtros de dados possíveis ao conceder permissões na tabela a uma entidade principal. Por exemplo, você pode criar um segundo filtro de dados para a tabela `orders`, que é um filtro de dados somente de segurança de linha. Referindo-se à captura de tela anterior, você pode escolher a opção Acesso a todas as colunas e incluir uma expressão de filtro de linha de `product_type<>pharma`. O nome desse filtro de dados pode ser `no-pharma`. Ele restringe o acesso a todas as linhas que têm a coluna `product_type` definida como “pharma”.

O objeto de solicitação para a operação `CreateDataCellsFilter` da API desse filtro de dados é o seguinte.

```
{
     "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"]
}
```

Em seguida, você poderia conceder SELECT na tabela `orders` com o filtro de dados `restrict-pharma` a um usuário administrativo, e SELECT na tabela `orders` com o filtro de dados `no-pharma` a usuários não administrativos. Para usuários do setor de saúde, você concederia SELECT na tabela `orders` com acesso total a todas as linhas e colunas (sem filtro de dados), ou talvez com outro filtro de dados que restringe o acesso às informações de preços.

É possível incluir ou excluir colunas aninhadas ao especificar a segurança em nível de coluna e em nível de linha em um filtro de dados. No exemplo a seguir, o acesso ao campo `product.offer` é especificado usando nomes de colunas qualificados (entre aspas duplas). Isso é importante para campos aninhados, a fim de evitar a ocorrência de erros quando os nomes das colunas contêm caracteres especiais e para manter a compatibilidade com versões anteriores das definições de segurança em nível de coluna de nível superior.

```
{
     "Name": "example_dcf",
     "DatabaseName": "example_db",
     "TableName": "example_table",
     "TableCatalogId": "111122223333",
     "RowFilter": { "FilterExpression": "customer.customerName <> 'John'" },
     "ColumnNames": ["customer", "\"product\".\"offer\""]
}
```

Consulte também

- Como gerenciar filtros de dados

Suporte PartiQL em expressões de filtro de linha

Você pode estruturar expressões de filtro de linha usando um subconjunto de tipos de dados, operadores e agregações do PartiQL. O Lake Formation não permite nenhum perfil PartiQL padrão ou definida pelo usuário na expressão de filtro. Você pode usar operadores de comparação para comparar colunas com constantes (por exemplo, `views >= 10000`), mas não pode comparar colunas com outras colunas.

Uma expressão de filtro de linha pode ser uma expressão simples ou uma expressão composta. O tamanho total da expressão deve ter menos de 2048 caracteres.

Expressões simples

Uma expressão simples terá o formato: `<column name > <comparison operator ><value >`

- Nome da coluna

  Deve ser uma coluna de dados de nível superior, uma coluna de partição ou uma coluna aninhada presente no esquema da tabela e deve pertencer aos Tipos de dados compatíveis listados abaixo.
- Operador de comparação

  Os seguintes operadores são compatíveis: `=, >, <, >=, <=, <>,!=, BETWEEN, IN, LIKE, NOT, IS [NOT] NULL`
- Todas as comparações de strings e correspondências de padrão `LIKE` diferenciam maiúsculas e minúsculas. Não é possível usar o operador IS [NOT] NULL em colunas de partição.
- Valor da coluna

  O valor da coluna deve corresponder ao tipo de dados do nome da coluna.

Expressão composta

Uma expressão composta terá o formato: `( <simple expression >) <AND/OR >(<simple expression >)`. Expressões compostas podem ser combinadas ainda mais usando operadores lógicos `AND/OR`.

Tipos de dados compatíveis

Filtros de linha que se referem a uma AWS Glue Data Catalog tabela que contém tipos de dados incompatíveis resultarão em um erro. A seguir estão os tipos de dados compatíveis com colunas e constantes da tabela, que são mapeados para tipos de Amazon Redshift dados:

- `STRING, CHAR, VARCHAR`
- `INT, LONG, BIGINT, FLOAT, DECIMAL, DOUBLE`
- `BOOLEAN`
- `STRUCT`

Para obter mais informações sobre tipos de dados no Amazon Redshift, consulte Tipos de dados no Guia do desenvolvedor de banco de dados do Amazon Redshift.

Expressões de filtro em linha

Example

Veja a seguir exemplos de expressões de filtro de linha válidas para uma tabela com colunas: `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

Veja a seguir exemplos de expressões de filtro de linha válidas para uma tabela com colunas aninhadas: `year > 2010 and customer.customerId <> 1`

Os campos aninhados em colunas de partição não devem ser referenciados ao definir expressões aninhadas em nível de linha.

As constantes de string devem estar entre aspas simples.

Palavras-chave reservadas

Se sua expressão de filtro de linha contiver palavras-chave PartiQL, você receberá um erro de análise, pois os nomes das colunas podem entrar em conflito com as palavras-chave. Quando isso acontecer, substitua os nomes das colunas por aspas duplas. Alguns exemplos de palavras-chave reservadas são “primeiro”, “último”, “asc”, “ausente”. Consulte a especificação do PartiQL para obter uma lista de palavras-chave reservadas.

Referência do PartiQL

Para obter mais informações sobre o PartiQL, consulte https://partiql.org/.

Permissões necessárias para consultar tabelas com filtragem em nível de célula

As permissões a seguir AWS Identity and Access Management (IAM) são necessárias para executar consultas em tabelas com filtragem em nível de célula.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:StartQueryPlanning",
                "lakeformation:GetQueryState",
                "lakeformation:GetWorkUnits",
                "lakeformation:GetWorkUnitResults"
            ],
            "Resource": "*"
        }
    ]
}
```

Para obter mais informações sobre as permissões de Lake Formation, consulte Referência de personas e permissões do IAM do Lake Formation.

Como gerenciar filtros de dados

Para implementar a segurança por coluna, por linha e por célula, é possível criar e manter filtros de dados. Cada filtro de dados pertence a uma tabela do catálogo de dados. Você pode criar vários filtros de dados para uma tabela e, em seguida, usar um ou mais deles ao conceder permissões na tabela. Também é possível definir e aplicar filtros de dados em colunas aninhadas que têm tipos de dados `struct` que permitem aos usuários acessar somente subestruturas de colunas aninhadas.

Você precisa da permissão `SELECT` com a opção de concessão para criar ou visualizar um filtro de dados. Para permitir que as entidades principais da sua conta visualizem e usem um filtro de dados, você pode conceder a permissão `DESCRIBE` sobre ele.

> **Note**
>
> O Lake Formation não dá suporte à concessão da permissão `Describe` em um filtro de dados, que é compartilhado de outra conta.

Você pode gerenciar filtros de dados usando o AWS Lake Formation console, a API ou o AWS Command Line Interface (AWS CLI).

Para obter informações sobre filtros de dados, consulte Filtros de dados no Lake Formation

Criar um filtro de dados

Você pode criar um ou mais filtros de dados para cada tabela do catálogo de dados.

Para criar um filtro de dados para uma tabela do catálogo de dados (console)

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

   Assine como administrador do data lake, proprietário da tabela de destino ou entidade principal que tenha uma permissão do Lake Formation na tabela de destino.
2. No painel de navegação, em catálogo de dados, escolha Filtros de dados.
3. Na página Filtros de dados, escolha Criar novo filtro.
4. Na caixa de diálogo Criar filtro de dados, insira as seguintes informações:

   - Nome do filtro de dados
   - Banco de dados de destino – especifique o banco de dados que contém a tabela.
   - Tabela de destino
   - Column-level acesso — deixe essa opção definida como Acesso a todas as colunas para especificar somente a filtragem de linhas. Escolha Incluir colunas ou Excluir colunas para especificar a filtragem de colunas ou células e, em seguida, especifique as colunas a serem incluídas ou excluídas.

     Colunas aninhadas: se você estiver aplicando o filtro em uma tabela que contenha colunas aninhadas, poderá especificar explicitamente as subestruturas das colunas da estrutura aninhada em um filtro de dados.

     Ao conceder a permissão SELECT a uma entidade principal nesse arquivador, a entidade principal que executa a consulta a seguir verá apenas os dados de `customer.customerName` e não de `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'
```

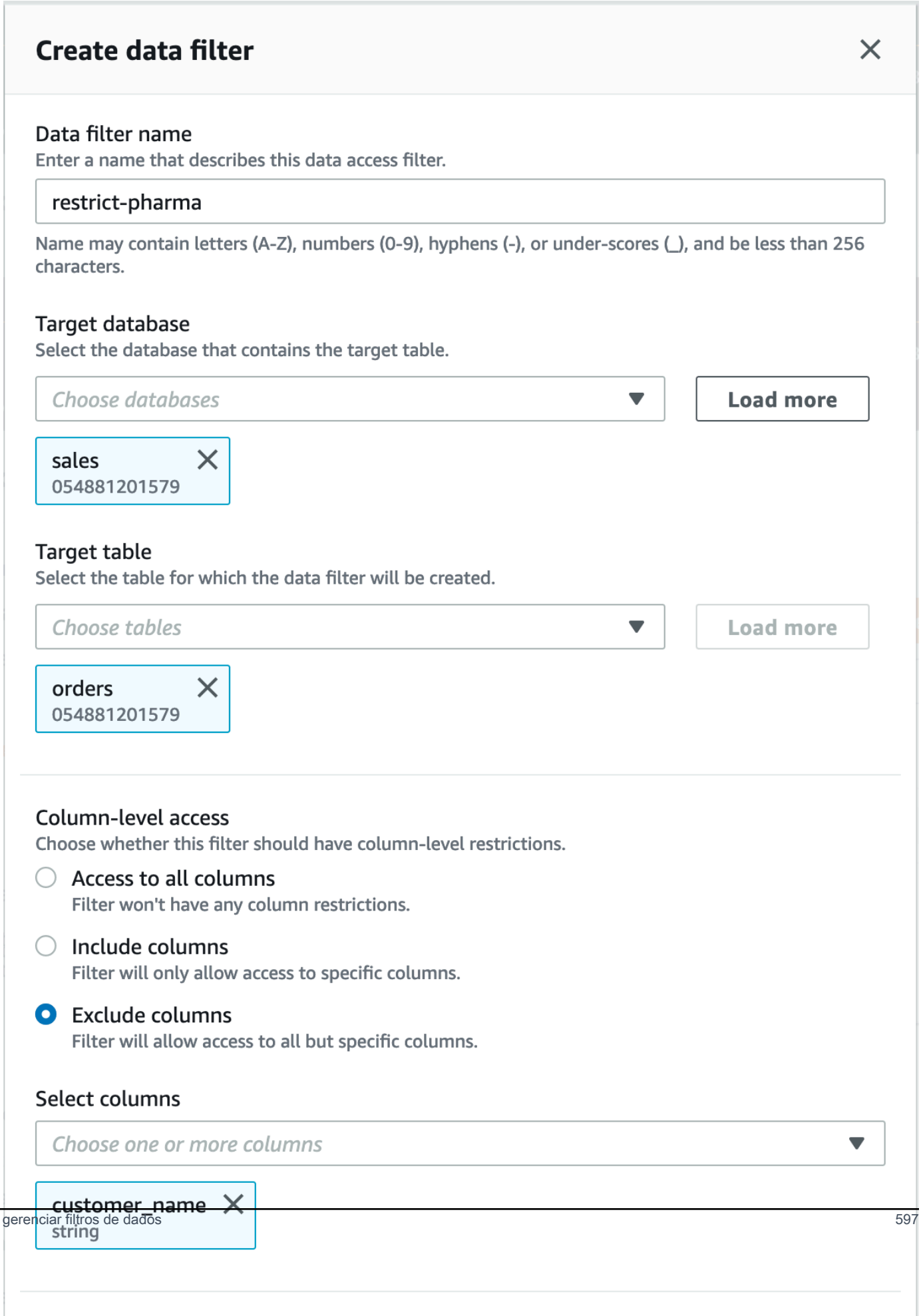
Ao conceder permissões à coluna `customer`, a entidade principal recebe o acesso à coluna e aos campos aninhados abaixo da coluna (`customerName` e `customerID`).

- Expressão de filtro de linha – insira uma expressão de filtro para especificar a filtragem de linha ou célula. Para obter os tipos de dados e operadores compatíveis, consulte Suporte PartiQL em expressões de filtro de linha. Selecione Acesso a todas as linhas para conceder acesso a todos.

  É possível incluir estruturas de coluna parciais de colunas aninhadas em uma expressão de filtro de linha para filtrar linhas que contenham valores específicos.

  Quando um principal recebe permissões para uma tabela com uma expressão `Select * from example_nestedtable where customer.customerName <>'John'` de filtro de linha e o Column-levelacesso é definido como Acesso a todas as colunas, os resultados da consulta mostram somente as linhas em que `customerName <>'John'` a avaliação é verdadeira.

A captura de tela a seguir mostra um filtro de dados que implementa a filtragem de células. Nas consultas com base na tabela `orders`, ela nega o acesso à coluna `customer_name` e mostra somente as linhas que têm “pharma” na coluna `product_type`.


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. Escolha Criar filtro.

Como criar um filtro de dados com políticas de filtro de células em um campo aninhado

Esta seção usa o seguinte exemplo de esquema para mostrar como criar um filtro de células de dados:

```
[
    { 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. Na página Criar um filtro de dados, insira um nome para o filtro de dados.
2. Depois, use o menu suspenso para escolher o nome do banco de dados e o nome da tabela.
3. Na seção de Column-level acesso, escolha Colunas incluídas e selecione uma coluna aninhada (`customer.customerName`).
4. Na seção de Row-level acesso, escolha a opção Acesso a todas as linhas.
5. Escolha Criar filtro.

   Ao conceder a permissão `SELECT` nesse filtro, a entidade principal obtém acesso a todas as linhas na coluna `customerName`.
6. Em seguida, defina outro filtro de dados para o mesmo database/table.
7. Na seção de Column-level acesso, escolha Colunas incluídas e selecione outra coluna aninhada (`customer.customerid`).
8. Na seção de Row-level acesso, escolha Filtrar linhas e insira uma expressão de filtro de linha (`customer.customerid <> 5`).
9. Escolha Criar filtro.

   Ao conceder a permissão `SELECT` nesse filtro, a entidade principal recebe acesso a todas as linhas no `customerName` e aos campos `customerId`, exceto à célula em que o valor é 5 na coluna `customerId`.

Conceder permissões de filtro de dados

Você pode conceder as permissões `SELECT`, `DESCRIBE` e `DROP` do Lake Formation sobre filtros de dados às entidades principais.

No início, somente você pode visualizar os filtros de dados que você cria para uma tabela. Para permitir que outra entidade principal visualize um filtro de dados e conceda permissões do catálogo de dados com o filtro de dados, você deve:

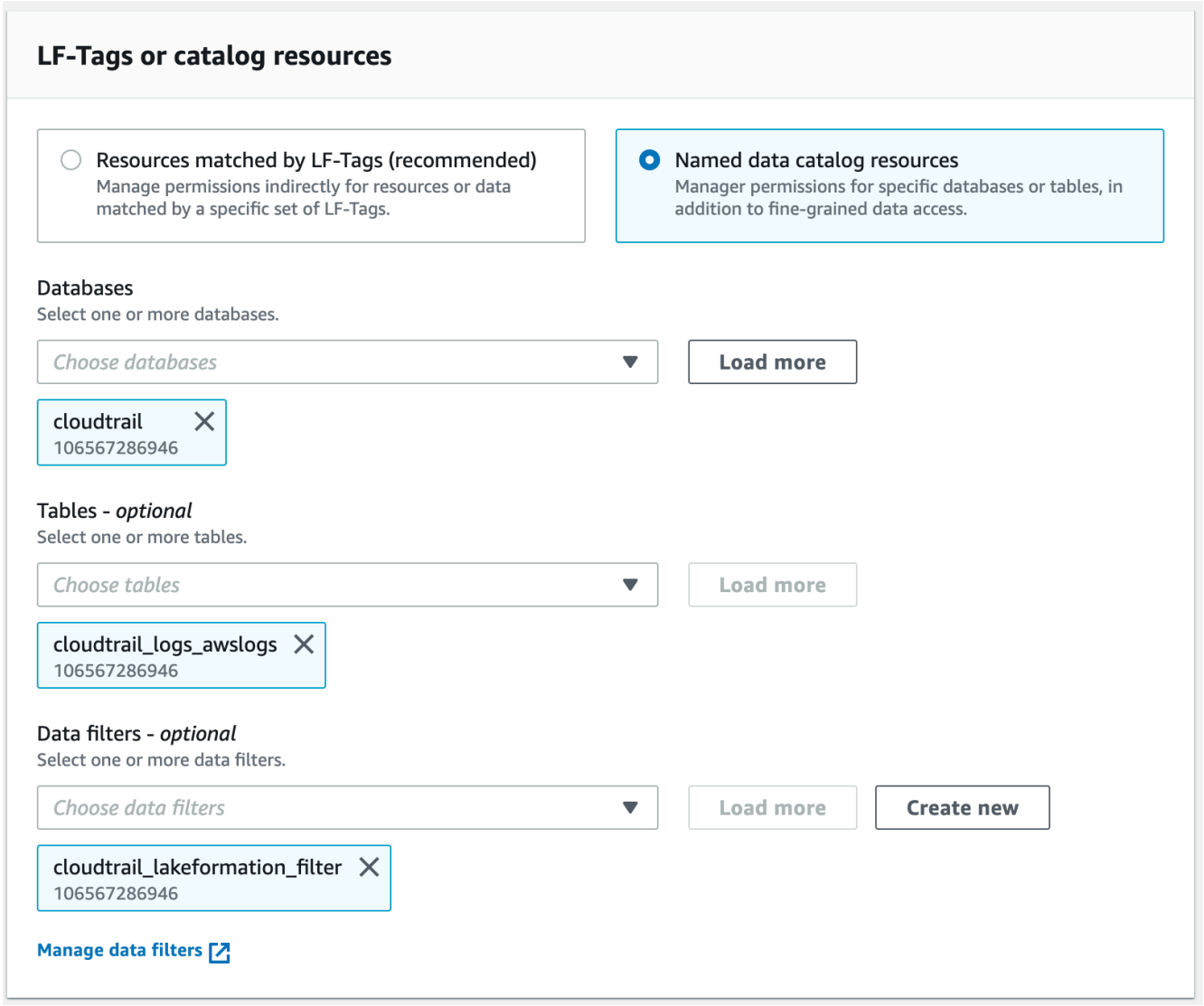
- Conceder `SELECT` uma tabela à entidade principal com a opção de concessão e aplique o filtro de dados à concessão.
- Conceder a permissão `DESCRIBE` ou `DROP` no filtro de dados da entidade principal.

Você pode conceder a `SELECT` permissão a uma AWS conta externa. Um administrador do data lake nessa conta pode então conceder essa permissão a outras entidades principais da conta. Ao conceder a uma conta externa, você deve incluir a opção de concessão para que o administrador da conta externa possa transmitir ainda mais a permissão para outros usuários na his/her conta. Ao conceder a uma entidade principal em sua conta, a concessão com a opção de concessão é opcional.

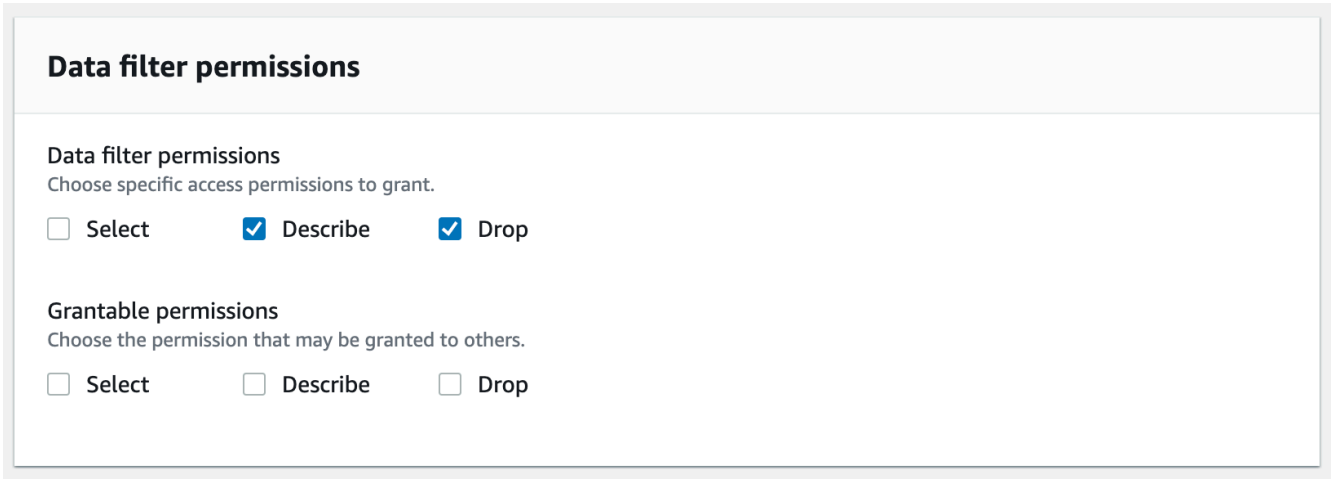
Você pode conceder e revogar permissões em filtros de dados usando o AWS Lake Formation console, a API ou o AWS Command Line Interface (AWS CLI).

Console

1. Faça login Console de gerenciamento da AWS e abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.
2. No painel de navegação, em Permissões, escolha Permissões do data lake.
3. Na página Permissões, na seção Permissões de dados, escolha Conceder.
4. Na página Conceder permissões de dados, escolha as entidades principais aos quais conceder as permissões.
5. Na seção LF-Tags ou recursos do catálogo, escolha Recursos do catálogo de dados nomeados. Em seguida, escolha o banco de dados, a tabela e o filtro de dados para os quais você deseja conceder permissões.



6. Na seção Permissões do filtro de dados, escolha as permissões que você deseja conceder às entidades principais selecionadas.



AWS CLI

- Insira um comando `grant-permissions`. Especifique `DataCellsFilter` para o argumento `resource` e especifique `DESCRIBE` ou `DROP` para o argumento `Permissions` e, opcionalmente, para o argumento `PermissionsWithGrantOption`.

  O exemplo a seguir concede a `DESCRIBE` a opção de concessão ao usuário `datalake_user1` no filtro de dados `restrict-pharma`, que pertence à tabela `orders` no banco de dados `sales` na conta AWS 1111-2222-3333.

  ```
  aws lakeformation grant-permissions --cli-input-json file://grant-params.json
  ```

  A seguir está o conteúdo do arquivo `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"]
  }
  ```

Conceder permissões de dados fornecidas por filtros de dados

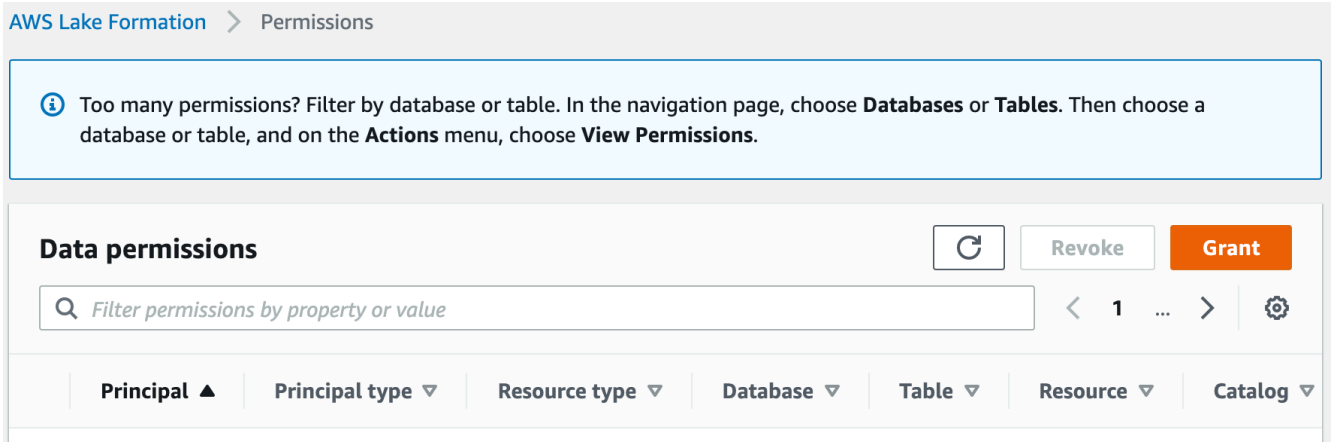
Os filtros de dados representam um subconjunto de dados em uma tabela. Para fornecer acesso aos dados às entidades principais, as permissões `SELECT` precisam ser concedidas a essas entidades principais. Com essa permissão, as entidades principais podem:

- Veja o nome real da tabela na lista de tabelas compartilhadas com suas contas.
- Crie filtros de dados na tabela compartilhada e conceda permissões a seus usuários nesses filtros de dados.

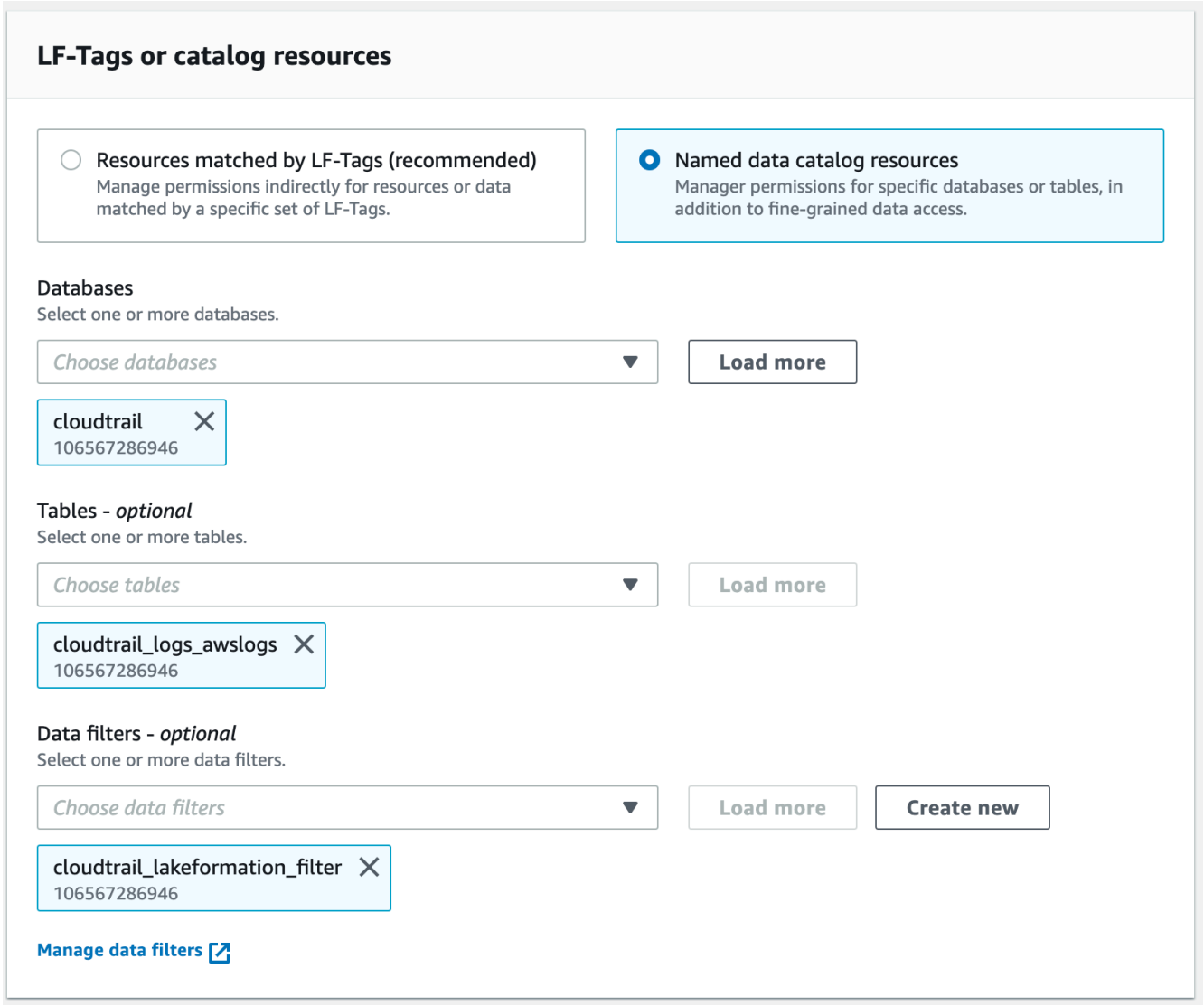
Console

Para conceder permissões SELECIONAR

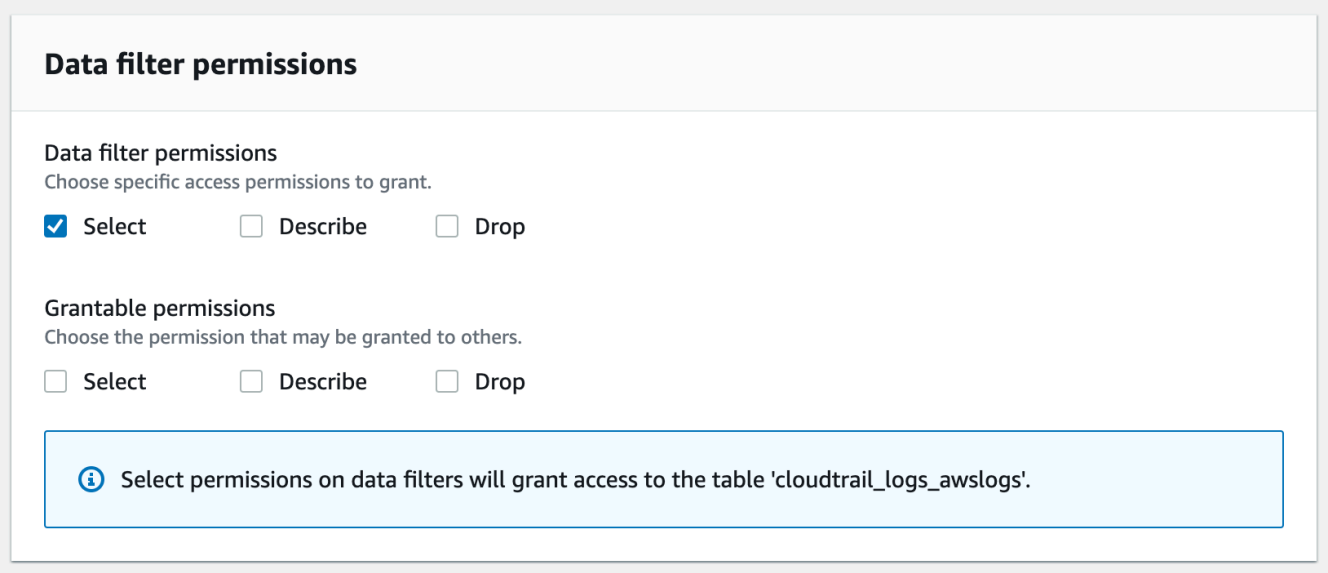
1. Acesse a página Permissões no console do Lake Formation e escolha Conceder.



2. Selecione as entidades principais às quais você deseja fornecer acesso e selecione Recursos do catálogo de dados nomeados.



3. Para fornecer acesso aos dados que o filtro representa, escolha Selecionar em Permissões do filtro de dados.



CLI

Insira um comando `grant-permissions`. Especifique `DataCellsFilter` para o argumento do recurso e especifique `SELECT` para o argumento Permissões.

O exemplo a seguir `SELECT` concede a opção de concessão ao usuário `datalake_user1` no filtro de dados`restrict-pharma`, que pertence à `orders` tabela no `sales` banco de dados em Conta da AWS `1111-2222-3333`.

```
aws lakeformation grant-permissions --cli-input-json file://grant-params.json
```

A seguir está o conteúdo do arquivo `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"]
}
```

Visualizando filtros de dados

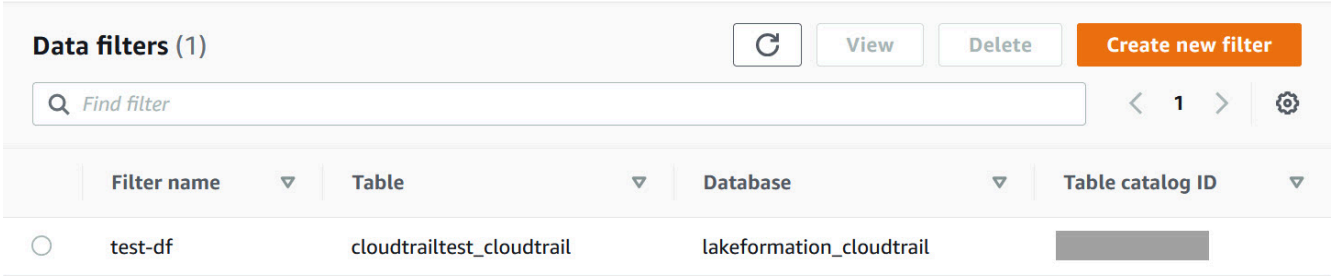
Você pode usar o console do Lake Formation ou AWS CLI a API do Lake Formation para visualizar os filtros de dados.

Para visualizar os filtros de dados, você deve ser administrador do data lake ou ter as permissões necessárias nos filtros de dados.

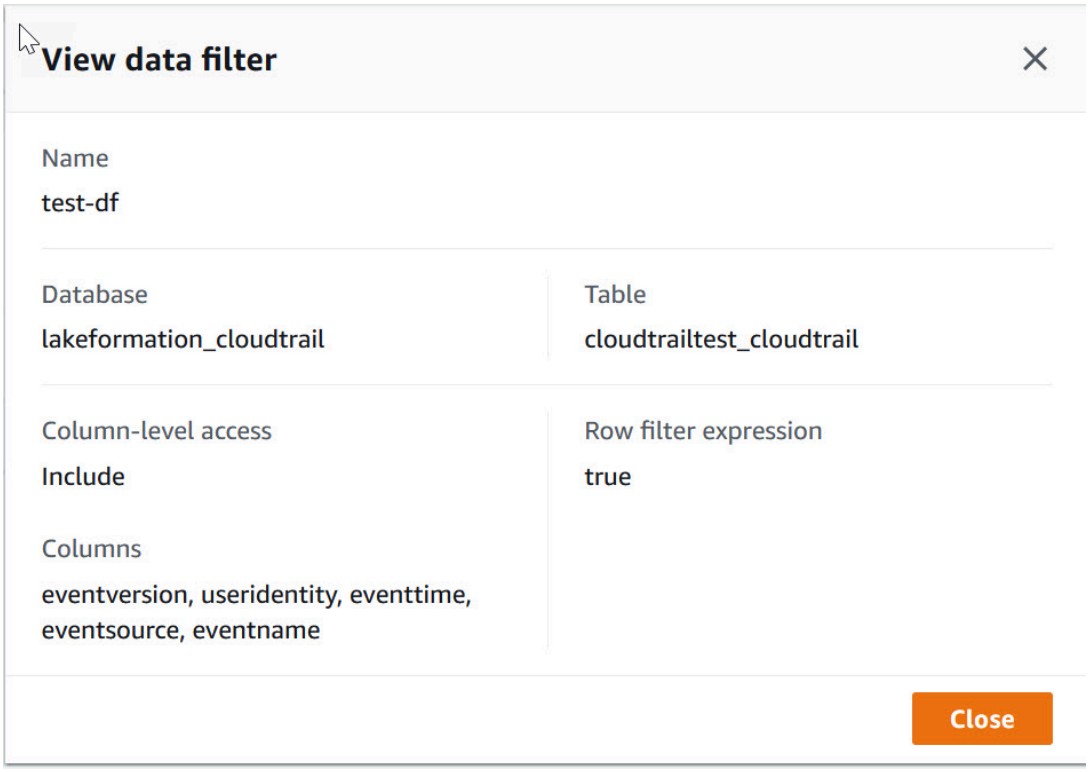
Console

1. Faça login Console de gerenciamento da AWS e abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.
2. No painel de navegação, em catálogo de dados, escolha Filtros de dados.

   A página exibe os filtros de dados aos quais você tem acesso.



3. Para visualizar os detalhes do filtro de dados, escolha o filtro de dados e, em seguida, escolha Exibir. Uma nova janela aparece com as informações detalhadas do filtro de dados.



AWS CLI

Insira um comando `list-data-cells-filter` e especifique um recurso de tabela.

O exemplo a seguir lista os filtros de dados para a tabela `cloudtrailtest_cloudtrail`.

```
aws lakeformation list-data-cells-filter --table '{ "CatalogId":"123456789012",
"DatabaseName":"lakeformation_cloudtrail", "Name":"cloudtrailtest_cloudtrail"}'
```

API/SDK

Use a API `ListDataCellsFilter` e especifique um recurso de tabela.

O exemplo a seguir usa Python para listar os 20 primeiros filtros de dados da tabela `myTable`.

```
response = client.list_data_cells_filter(
    Table = {
        'CatalogId': '111122223333',
        'DatabaseName': 'mydb',
        'Name': 'myTable'
    },
```

```
    MaxResults=20
)
```

Listando permissões de filtro de dados

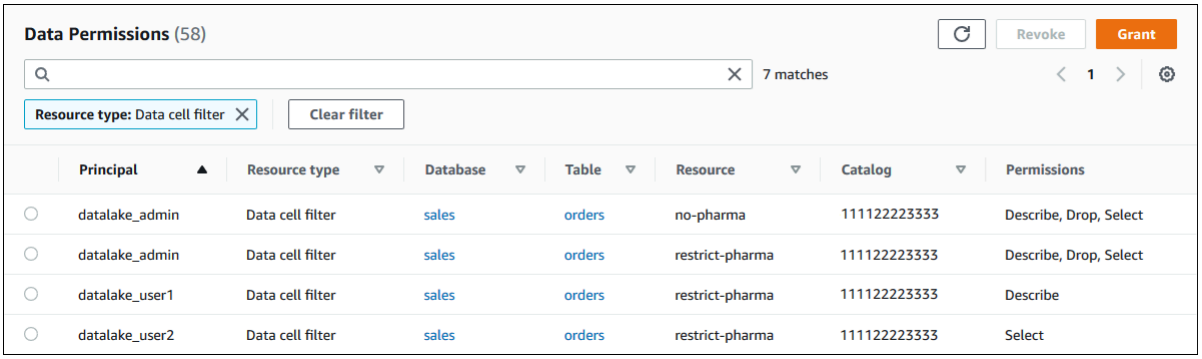
Você pode usar o console do Lake Formation para visualizar as permissões concedidas nos filtros de dados.

Para ver as permissões em um filtro de dados, você deve ser administrador do data lake ou ter as permissões necessárias no filtro de dados.

Console

1. Faça login Console de gerenciamento da AWS e abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.
2. No painel de navegação, em Permissões, escolha Permissões de dados.
3. Na página Permissões de dados, clique ou toque no campo de pesquisa e, no menu Propriedades, escolha Tipo de recurso.
4. No menu Tipo de recurso, escolha Tipo de recurso: filtro de células de dados.

   Os filtros de dados nos quais você tem permissões estão listados. Talvez seja necessário rolar horizontalmente para ver as colunas de Permissões e de Concessão.



AWS CLI

- Insira um comando `list-permissions`. Especifique `DataCellsFilter` para o argumento `resource` e especifique `DESCRIBE` ou `DROP` para o argumento `Permissions` e, opcionalmente, para o argumento `PermissionsWithGrantOption`.

O exemplo a seguir lista as permissões `DESCRIBE` com a opção de concessão no filtro de dados `restrict-pharma`. Os resultados são limitados às permissões concedidas para o diretor `datalake_user1` e a `orders` tabela no `sales` banco de dados na AWS conta 1111-2222-3333.

```
aws lakeformation list-permissions --cli-input-json file://list-params.json
```

A seguir está o conteúdo do arquivo `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"]
}
```

Visualizar permissões de banco de dados e tabelas no Lake Formation

Você pode visualizar as permissões do Lake Formation concedidas em um banco de dados ou uma tabela do catálogo de dados. Você pode fazer isso usando o console do Lake Formation, a API ou o AWS Command Line Interface (AWS CLI).

Ao usar o console, você pode visualizar as permissões nas páginas Bancos de dados, Tabelas ou Permissões de dados.

> **Note**
>
> Se você não for administrador de banco de dados ou proprietário de recursos, poderá ver as permissões que outras entidades principais têm sobre o recurso somente se tiver uma permissão do Lake Formation sobre o recurso com a opção de concessão.
> Além das permissões necessárias do Lake Formation, você precisa das permissões AWS Identity and Access Management (IAM) `glue:GetDatabases` `glue:GetDatabaseglue:GetTables`,`glue:GetTable`,, `lakeformation:ListPermissions` e.

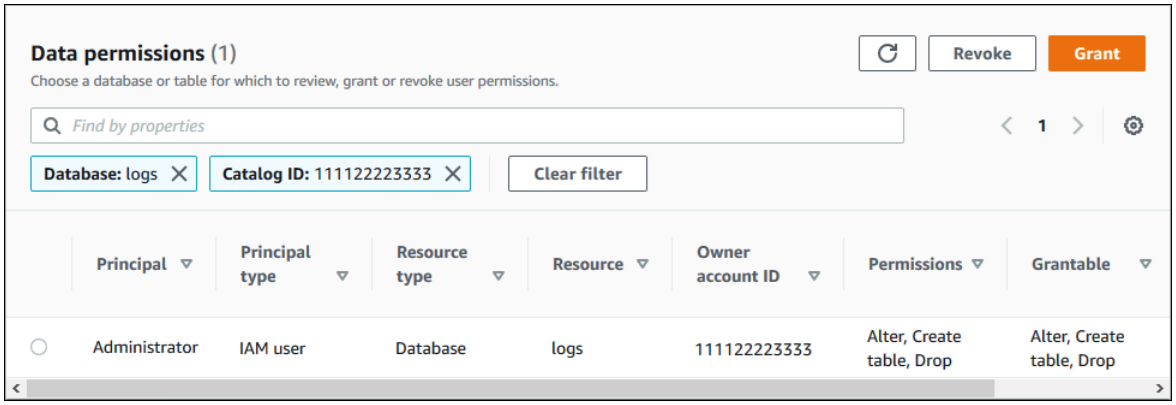
Para visualizar as permissões em um banco de dados (console, a partir da página Bancos de dados)

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

   Faça login como administrador do data lake, criador do banco de dados ou como usuário que tenha qualquer permissão do Lake Formation no banco de dados com a opção de concessão.
2. No painel de navegação, escolha Bancos de dados.
3. Escolha um banco de dados e, no menu Ações, escolha Exibir permissões.

   > **Note**
   >
   > Se você escolher um link de recurso de banco de dados, o Lake Formation exibirá as permissões no link do recurso, não no banco de dados de destino do link do recurso.

   A página Permissões de dados lista todas as permissões do Lake Formation para o banco de dados. O nome do banco de dados e o ID do catálogo (ID da AWS conta) do proprietário do banco de dados aparecem como rótulos na caixa de pesquisa. Os mosaicos indicam que um filtro foi aplicado para listar permissões somente para esse banco de dados. Você pode ajustar o filtro fechando um quadro ou escolhendo Limpar filtro.



Para visualizar as permissões em um banco de dados (console, a partir da página Permissões de dados)

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

   Faça login como administrador do data lake, criador do banco de dados ou como usuário que tenha qualquer permissão do Lake Formation no banco de dados com a opção de concessão.
2. No painel de navegação, selecione Permissões de dados.
3. Posicione o cursor na caixa de pesquisa e, no menu Propriedades exibido, escolha Banco de dados.
4. No menu de bancos de dados exibido, escolha um banco de dados.

> **Note**
>
> Se você escolher um link de recurso de banco de dados, o Lake Formation exibirá as permissões no link do recurso, não no banco de dados de destino do link do recurso.

   A página Permissões de dados lista todas as permissões do Lake Formation para o banco de dados. O nome do banco de dados aparece como um quadro abaixo da caixa de pesquisa. Os blocos indicam que um filtro foi aplicado para listar permissões somente para esse banco de dados. Você pode remover o filtro fechando o quadro ou escolhendo Limpar filtro.

Para visualizar as permissões em uma tabela (console, a partir da página Tabelas)

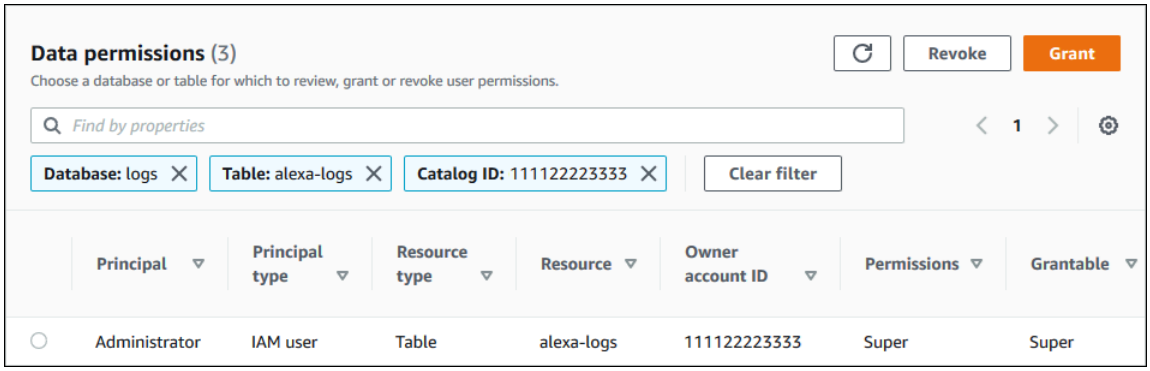
1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

Faça login como administrador do data lake, criador da tabela ou como usuário que tenha qualquer permissão do Lake Formation na tabela com a opção de concessão.

2. No painel de navegação, selecione Tabelas.
3. Escolha uma tabela e, no menu Ações, selecione Exibir permissões.

> **Note**
>
> Se você escolher um link de recurso de tabela, o Lake Formation exibirá as permissões no link do recurso, não na tabela de destino do link do recurso.

A página Permissões de dados lista todas as permissões do Lake Formation para a tabela. O nome da tabela, o nome do banco de dados que contém a tabela e a ID do catálogo (ID da AWS conta) do proprietário da tabela aparecem como rótulos na caixa de pesquisa. As etiquetas indicam que um filtro foi aplicado para listar permissões apenas para essa tabela. Você pode ajustar o filtro fechando uma etiqueta ou escolhendo Limpar filtro.



Para visualizar as permissões em uma tabela (console, a partir da página Permissões de dados)

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

   Faça login como administrador do data lake, criador do banco de dados ou como usuário que tenha qualquer permissão do Lake Formation no banco de dados com a opção de concessão.
2. No painel de navegação, selecione Permissões de dados.
3. Posicione o cursor na caixa de pesquisa e, no menu Propriedades exibido, escolha Banco de dados.
4. No menu de Bancos de dados exibido, escolha um banco de dados.

> **Important**
>
> Se você quiser ver as permissões em uma tabela que foi compartilhada com sua AWS conta a partir de uma conta externa, você deve escolher o banco de dados na conta externa que contém a tabela, não um link de recurso para o banco de dados.

A página Permissões de dados lista todas as permissões do Lake Formation para o banco de dados.

5. Posicione o cursor na caixa de pesquisa na parte superior da página e, no menu Propriedades exibido, escolha Tabelas.
6. No menu Tabelas exibido, escolha uma tabela.

   A página Permissões de dados lista todas as permissões do Lake Formation para o banco de dados. O nome da tabela e o nome do banco de dados que contém a tabela aparecem como blocos sob a caixa de pesquisa. Os mosaicos indicam que um filtro foi aplicado para listar permissões somente para essa tabela. Você pode ajustar o filtro fechando um quadro ou escolhendo Limpar filtro.

Para ver as permissões em uma tabela (AWS CLI)

- Insira um comando `list-permissions`.

  O exemplo a seguir lista as permissões em uma tabela compartilhada de uma conta externa. A `CatalogId` propriedade é o ID da AWS conta externa, e o nome do banco de dados se refere ao banco de dados na conta externa que contém a tabela.

```
aws lakeformation list-permissions  --resource-type TABLE --resource '{ "Table":
 {"DatabaseName":"logs", "Name":"alexa-logs", "CatalogId":"123456789012"}}'
```

Revogando a permissão usando o console Lake Formation

Você pode usar o console para revogar todos os tipos de permissões do Lake Formation – permissões de catálogo de dados, permissões de etiquetas de políticas, permissões de filtro de dados e permissões de localização.

Para revogar as permissões do Lake Formation em um recurso (console)

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

   Faça login como administrador do data lake ou como usuário que recebeu permissões com a opção de concessão no recurso.
2. No painel de navegação, em Permissões, escolha Permissões LF-Tags e permissões do Data Lake ou Localizações de dados.
3. Selecione a permissão ou o local e, em seguida, escolha Revogar.
4. Na caixa de diálogo exibida, escolha Revogar.

Cross-account compartilhamento de dados em Lake Formation

Os recursos de várias contas do Lake Formation permitem que os usuários compartilhem com segurança lagos de dados distribuídos em várias AWS organizações ou diretamente com os diretores do IAM em outra conta Contas da AWS, fornecendo acesso refinado aos metadados do Catálogo de Dados e aos dados subjacentes. As grandes empresas geralmente usam várias Contas da AWS, e muitas dessas contas podem precisar acessar um data lake gerenciado por uma única Conta da AWS. Os usuários e as tarefas de AWS Glue extração, transformação e carregamento (ETL) podem consultar e unir tabelas em várias contas e ainda aproveitar as proteções de dados em nível de tabela e coluna do Lake Formation.

Quando você concede permissões do Lake Formation em um recurso do Catálogo de Dados para uma conta externa ou diretamente para um diretor do IAM em outra conta, o Lake Formation usa o serviço AWS Resource Access Manager (AWS RAM) para compartilhar o recurso. Se a conta do concedido estiver na mesma organização da conta do concedente, o recurso compartilhado estará disponível imediatamente para o concedido. Se a conta do beneficiário não estiver na mesma organização, AWS RAM envia um convite à conta do beneficiário para aceitar ou rejeitar a concessão do recurso. Em seguida, para disponibilizar o recurso compartilhado, o administrador do data lake na conta do beneficiário deve usar o AWS RAM console ou AWS CLI aceitar o convite.

O Lake Formation permite o compartilhamento de recursos do catálogo de dados com contas externas no modo de acesso híbrido. O modo de acesso híbrido oferece a flexibilidade de habilitar seletivamente as permissões do Lake Formation para bancos de dados e tabelas no seu AWS Glue Data Catalog.

Com o modo de acesso híbrido, agora você tem um caminho incremental que permite definir permissões do Lake Formation para um conjunto específico de usuários sem interromper as políticas de permissão de outros usuários ou workloads existentes.

Para obter mais informações, consulte Modo de acesso híbrido.

Compartilhamento direto entre contas

As entidades principais autorizadas podem compartilhar recursos explicitamente com uma entidade principal do IAM em uma conta externa. Esse atributo é útil quando o proprietário da conta deseja ter controle sobre quem na conta externa pode acessar os recursos. As permissões que a entidade principal do IAM receberá serão uma união de concessões diretas e concessões em nível de conta que serão transferidas em cascata para as entidades principais. Somente o destinatário da permissão pode visualizar as concessões diretas entre contas. A entidade principal que recebe o compartilhamento de recursos não pode compartilhar o recurso com outras entidades principais.

Métodos para compartilhar recursos do catálogo de dados

Com uma única operação de concessão do Lake Formation, você pode conceder permissões entre contas nos seguintes recursos do catálogo de dados.

- Um banco de dados
- Uma tabela individual (com filtragem de coluna opcional)
- Algumas tabelas selecionadas
- Todas as tabelas em um banco de dados (usando o curinga Todas as Tabelas)

Há duas opções para compartilhar seus bancos de dados e tabelas com outra conta Conta da AWS ou com diretores do IAM em outra conta.

- Controle de acesso baseado em tags Lake Formation (LF-TBAC) (recomendado)

  O controle de acesso baseado em tags do Lake Formation é uma estratégia de autorização que define permissões com base em atributos. Você pode usar o controle de acesso baseado em tags para compartilhar recursos do Catálogo de Dados (bancos de dados, tabelas e colunas) com diretores externos do IAM, Contas da AWS Organizations and Organizational Units (OUs). Em Lake Formation, esses atributos são chamados LF-tags. Para obter mais informações, consulte Gerenciamento de um data lake usando o controle de acesso baseado em tags do Lake Formation.

> **Note**
>
> O LF-TBAC método de concessão de permissões do Catálogo de Dados é usado AWS Resource Access Manager para concessões entre contas.
> O Lake Formation agora oferece suporte à concessão de permissões entre contas para organizações e unidades organizacionais usando o LF-TBAC método.
> Para habilitar esse recurso, você precisa atualizar as Configurações de versão entre contas para a Versão 3 ou posterior.
> Para obter mais informações, consulte Como atualizar as configurações da versão de compartilhamento de dados entre contas.

- Recursos nomeados do Lake Formation

  O compartilhamento de dados entre contas do Lake Formation usando o método de recurso nomeado permite que você conceda permissões do Lake Formation com uma opção de concessão em tabelas e bancos de dados do Catálogo de Dados para entidades externas Contas da AWS, diretores do IAM, organizações ou unidades organizacionais. A operação de concessão compartilha automaticamente esses recursos.

> **Note**
>
> Você também pode permitir que o AWS Glue rastreador acesse um armazenamento de dados em uma conta diferente usando as credenciais do Lake Formation. Para obter mais informações, consulte Cross-account rastreamento no Guia do AWS Glue desenvolvedor.

Serviços integrados, como o Athena e o Amazon Redshift Spectrum, exigem links de recursos para poder incluir recursos compartilhados nas consultas. Para obter mais informações sobre os links de recursos, consulte Como os links de recursos funcionam no Lake Formation.

Para conhecer as limitações e as considerações, consulte Práticas recomendadas e considerações sobre compartilhamento de dados entre contas.

Tópicos

- Pré-requisitos
- Como atualizar as configurações da versão de compartilhamento de dados entre contas

- Compartilhando tabelas e bancos de dados do Catálogo de Dados entre Contas da AWS ou diretores do IAM de contas externas
- Conceder permissões em um banco de dados ou tabela compartilhada com sua conta
- Como conceder permissões de links de recursos
- Como acessar os dados subjacentes de uma tabela compartilhada
- Cross-account CloudTrail registro
- Gerenciando permissões entre contas usando ambos AWS Glue e Lake Formation
- Visualizando todas as concessões entre contas usando a operação de GetResourceShares API

Tópicos relacionados

- Visão geral das permissões do Lake Formation
- Acessar e visualizar tabelas e bancos de dados compartilhados do catálogo de dados
- Criação de links de recursos
- Resolução de problemas de acesso entre contas

Pré-requisitos

Antes que sua AWS conta possa compartilhar recursos do Catálogo de Dados (catálogos, bancos de dados e tabelas) com outra conta ou diretores em outra conta, e antes que você possa acessar os recursos compartilhados com sua conta, os seguintes pré-requisitos devem ser atendidos.

Requisitos gerais de compartilhamento de dados entre contas

- Para compartilhar bancos de dados e tabelas do Data Catalog no modo de acesso híbrido e compartilhar objetos nos catálogos federados, você precisa atualizar as Configurações de versão entre contas para a Versão 4.
- Antes de conceder permissões entre contas em um recurso do catálogo de dados, você deve revogar todas as permissões do Lake Formation do grupo `IAMAllowedPrincipals` para o recurso. Se a entidade principal solicitante tiver permissões entre contas para acessar um recurso, e a permissão `IAMAllowedPrincipals` existir no recurso, Lake Formation exibirá `AccessDeniedException`.

Esse requisito é válido somente quando você registra a localização dos dados subjacentes no modo Lake Formation. Se você registrar a localização dos dados no modo híbrido, as permissões do grupo `IAMAllowedPrincipals` poderão existir no banco de dados ou na tabela compartilhada.

- Nos bancos de dados que contêm tabelas que você cogita compartilhar, é necessário evitar que novas tabelas tenham uma concessão padrão de `Super` para `IAMAllowedPrincipals`. No console do Lake Formation, edite o banco de dados e desative Usar somente o controle de acesso do IAM para novas tabelas nesse banco de dados ou insira o AWS CLI comando a seguir, `database` substituindo-o pelo nome do banco de dados. Se a localização dos dados subjacentes estiver registrada no modo de acesso híbrido, você não precisará alterar essa configuração padrão. No modo de acesso híbrido, o Lake Formation permite que você aplique seletivamente as permissões do Lake Formation e as políticas de permissões do IAM para o Amazon S3 AWS Glue e no mesmo recurso.

```
aws glue update-database --name database --database-input
 '{"Name":"database","CreateTableDefaultPermissions":[]}'
```

- Para conceder permissões entre contas, o concedente deve ter as permissões necessárias AWS Identity and Access Management (IAM) sobre AWS Glue o serviço. AWS RAM A política AWS gerenciada `AWSLakeFormationCrossAccountManager` concede as permissões necessárias.

  Os administradores de data lake em contas que recebem compartilhamentos de recursos usando AWS RAM devem ter a seguinte política adicional. Ele permite que o administrador aceite convites AWS RAM de compartilhamento de recursos. Permitir também que o administrador habilite o compartilhamento de recursos com organizações.

  JSON

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

```
                "ram:EnableSharingWithAwsOrganization"
            ],
            "Resource": "*"
        }
    ]
}
```

- Se você quiser compartilhar recursos do Catálogo de Dados com AWS Organizations ou com unidades organizacionais, o compartilhamento com organizações deve estar ativado em AWS RAM.

  Para obter informações sobre como ativar o compartilhamento com organizações, consulte Habilitar o compartilhamento com AWS organizações no Guia AWS RAM do usuário.

  Você deve ter a permissão `ram:EnableSharingWithAwsOrganization` para habilitar o compartilhamento com organizações.

- Para compartilhar recursos diretamente com uma entidade principal do IAM em outra conta, você precisa atualizar as Configurações Versão Entre Contas para a Versão 3. Essa configuração está disponível na página de Configurações do catálogo de dados. Se você estiver usando a Versão 1, consulte as instruções para atualizar a configuraçãoComo atualizar as configurações da versão de compartilhamento de dados entre contas.
- Você não pode compartilhar recursos do Catálogo de Dados criptografados com a chave gerenciada do AWS Glue serviço com outra conta. Você pode compartilhar somente recursos do catálogo de dados criptografados com a chave de criptografia do cliente, e a conta que recebe o compartilhamento de recursos deve ter permissões na chave de criptografia do catálogo de dados para descriptografar os objetos.

Cross-account compartilhamento de dados usando LF-TBAC requisitos

- Para compartilhar recursos do Catálogo de Dados com AWS Organizations unidades organizacionais (OUs), você precisa atualizar as configurações da versão da conta cruzada para a versão 3 ou superior.
- Para compartilhar os recursos do catálogo de dados com a versão 3 das Configurações de versão entre contas, o concedente precisa ter as permissões de IAM definidas na política `AWSLakeFormationCrossAccountManager` gerenciada por AWS em sua conta.
- Se você estiver usando a versão 1 ou a versão 2 das configurações de versão da conta cruzada, deverá ter uma política de recursos do Catálogo de Dados (`glue:PutResourcePolicy`) que

habilite LF-TBAC. Para obter mais informações, consulte Gerenciando permissões entre contas usando ambos AWS Glue e Lake Formation.

- Se você estiver usando atualmente uma política de recursos do Catálogo de AWS Glue Dados para compartilhar recursos e quiser conceder permissões entre contas usando a versão 3 ou superior das configurações da versão entre contas, deverá adicionar a `glue:ShareResource` permissão nas Configurações do Catálogo de Dados usando a operação de `glue:PutResourcePolicy` API, conforme mostrado na Gerenciando permissões entre contas usando ambos AWS Glue e Lake Formation seção. Essa política não é necessária se sua conta não fez concessões entre contas usando a política de recursos do catálogo de dados AWS Glue (permissão de uso `glue:PutResourcePolicy` da versão 1 e versão 2) para conceder acesso entre contas.

```
{
      "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"
      ]
    }
```

- Se sua conta fez compartilhamentos entre contas usando a política de recursos do Catálogo de AWS Glue Dados e você está usando o método de recurso nomeado ou LF-TBAC com as configurações de contas cruzadas versão 3 para compartilhar recursos, que usa AWS RAM para compartilhar recursos, você deve definir o `EnableHybrid` argumento para `'true'` quando invocar a operação da `glue:PutResourcePolicy` API. Para obter mais informações, consulte Gerenciando permissões entre contas usando ambos AWS Glue e Lake Formation.

Configuração necessária em cada conta que acessa o recurso compartilhado

- Se você estiver compartilhando recursos com Contas da AWS, pelo menos um usuário na conta do consumidor deve ser administrador do data lake para visualizar os recursos compartilhados.

Para obter informações sobre como criar um administrador de data lake, consulte Crie um administrador de data lake.

O administrador do data lake pode conceder permissões do Lake Formation sobre os recursos compartilhados com outras entidades principais da conta. Outras entidades principais não podem acessar recursos compartilhados até que o administrador do data lake conceda permissões sobre os recursos.

- Serviços integrados, como o Athena e o Redshift Spectrum, exigem links de recursos para poder incluir recursos compartilhados nas consultas. Entidades principais precisam criar um link de recurso em seu catálogo de dados para um recurso compartilhado de outra Conta da AWS. Para obter mais informações sobre os links de recursos, consulte Como os links de recursos funcionam no Lake Formation.
- Quando um recurso é compartilhado diretamente com uma entidade principal do IAM, para consultar a tabela usando o Athena, a entidade principal precisa criar um link de recurso. Para criar um link de recurso, a entidade principal precisa da permissão `CREATE_TABLE` ou `CREATE_DATABASE` do Lake Formation, e da permissão `glue:CreateTable` ou `glue:CreateDatabase` do IAM.

  Se a conta do produtor compartilhar uma tabela diferente no mesmo banco de dados com o mesmo ou outra entidade principal, esse principal poderá consultar imediatamente a tabela.

> **Note**
>
> Para o administrador do data lake e para as entidades principais às quais o administrador do data lake concedeu permissões, os recursos compartilhados aparecem no catálogo de dados como se fossem recursos locais (de propriedade). Tarefas de extração, transformação e carregamento (ETL) podem acessar os dados subjacentes dos recursos compartilhados. Para recursos compartilhados, as páginas Tabelas e bancos de dados no console do Lake Formation exibem o ID da conta do proprietário.
> Quando os dados subjacentes de um recurso compartilhado são acessados, os eventos de CloudTrail registro são gerados na conta do destinatário do recurso compartilhado e na conta do proprietário do recurso. Os CloudTrail eventos podem conter o ARN do principal que acessou os dados, mas somente se a conta do destinatário optar por incluir o ARN principal nos registros. Para obter mais informações, consulte Cross-account CloudTrail registro.

Como atualizar as configurações da versão de compartilhamento de dados entre contas

De tempos em tempos, AWS Lake Formation atualiza as configurações de compartilhamento de dados entre contas para distinguir as alterações feitas no AWS RAM uso e para oferecer suporte às atualizações feitas no recurso de compartilhamento de dados entre contas. Quando o Lake Formation faz isso, ele cria uma nova versão das Configurações de versão entre contas.

Principais diferenças entre as versões de Configurações Entre Contas

Para obter mais informações sobre como o compartilhamento de dados entre contas funciona em diferentes versões de Configurações de versão entre contas, consulte as seções a seguir.

> **Note**
>
> Para compartilhar dados com outra conta, o concedente deve ter permissões de política do IAM gerenciadas por `AWSLakeFormationCrossAccountManager`. Isso é um pré-requisito para todas as versões.
> A atualização da versão das Configurações de versão entre contas não afeta as permissões que o destinatário tem nos recursos compartilhados. Isso é aplicável ao atualizar da versão 1 para a versão 2, da versão 2 para a versão 3 e da versão 1 para a versão 3. Veja as considerações listadas abaixo ao atualizar as versões.

Versão 1

Método de recurso nomeado: mapeia cada concessão de permissão entre contas do Lake Formation para um compartilhamento AWS RAM de recursos. O usuário (no perfil de concedente ou entidade principal) não precisa de permissões adicionais.

LF-TBAC método: as permissões concedidas pelo Cross-account Lake Formation não são usadas AWS RAM para compartilhar dados. O usuário deve ter a permissão `glue:PutResourcePolicy`.

Benefícios da atualização de versões: Versão inicial — não aplicável.

Considerações ao atualizar versões: Versão inicial — não aplicável.

Versão 2

Método de recurso nomeado: otimiza o número de compartilhamentos de AWS RAM recursos mapeando várias concessões de permissão entre contas com um compartilhamento de AWS RAM recursos. Usuários não precisam de permissões adicionais.

LF-TBAC método: as permissões concedidas pelo Cross-account Lake Formation não são usadas AWS RAM para compartilhar dados. O usuário deve ter a permissão `glue:PutResourcePolicy`.

Benefícios da atualização de versões: configuração escalável entre contas por meio da utilização ideal da capacidade. AWS RAM

Considerações ao atualizar versões: os usuários que desejam conceder permissões entre contas do Lake Formation devem ter as permissões na política `AWSLakeFormationCrossAccountManager` AWS gerenciada. Caso contrário, você precisará ter permissões `ram:AssociateResourceShare` e `ram:DisassociateResourceShare` para compartilhar recursos com sucesso com outra conta.

Versão 3

Método de recurso nomeado: otimiza o número de compartilhamentos de AWS RAM recursos mapeando várias concessões de permissão entre contas com um compartilhamento de AWS RAM recursos. Usuários não precisam de permissões adicionais.

LF-TBAC método: o Lake Formation usa AWS RAM para doações entre contas. O usuário deve adicionar cola: ShareResource declaração à `glue:PutResourcePolicy` permissão. O destinatário deve aceitar convites de compartilhamento de recursos de. AWS RAM

Benefícios da atualização de versões: Suporta os seguintes recursos:

- Permite compartilhar recursos explicitamente com uma entidade principal do IAM em uma conta externa.

  Para obter mais informações, consulte Conceder permissões nos recursos do Catálogo de Dados.
- Permite compartilhamentos entre contas usando o LF-TBAC método para Organizações ou unidades organizacionais (OUs).
- Elimina a sobrecarga de manter AWS Glue políticas adicionais para subsídios entre contas.

Considerações ao atualizar versões: Quando você usa o LF-TBAC método para compartilhar recursos, se o concedente usa uma versão inferior à versão 3 e o destinatário está usando

a versão 3 ou superior, o concedente recebe a seguinte mensagem de erro: “Solicitação de concessão de contas cruzadas inválida. Consumer account has opt-in to cross account version: v3. `CrossAccountVersion`Atualize `DataLakeSetting` para a versão mínima v3 (Serviço: AmazonDataCatalog; Código de status: 400; Código de erro: InvalidInputException)”. No entanto, se o concedente usar a versão 3 e o destinatário estiver usando a versão 1 ou a versão 2, as concessões entre contas serão aprovadas com LF-Tags êxito.

Cross-account as concessões feitas usando o método de recurso nomeado são compatíveis em diferentes versões. Mesmo que a conta do concedente esteja usando uma versão mais antiga (versão 1 ou 2) e a conta do destinatário esteja usando uma versão mais recente (versão 3 ou posterior), a funcionalidade de acesso entre contas funciona perfeitamente, sem problemas ou erros de compatibilidade.

Para compartilhar recursos diretamente com as entidades principais do IAM em outra conta, somente o concedente precisa usar a versão 3.

Cross-account concessões feitas usando o LF-TBAC método exigem que os usuários tenham uma política de AWS Glue Data Catalog recursos na conta. Quando você atualiza para a versão 3, LF-TBAC concede usos AWS RAM. Para permitir que as concessões AWS RAM baseadas em várias contas sejam bem-sucedidas, você deve adicionar a `glue:ShareResource` declaração às suas políticas de recursos existentes do Catálogo de Dados, conforme mostrado na Gerenciando permissões entre contas usando ambos AWS Glue e Lake Formation seção.

Versão 4

O concedente precisa da versão 4 ou posterior para compartilhar os recursos do Data Catalog no modo de acesso híbrido ou compartilhar objetos em um catálogo federado.

Versão 5

A versão 5 da conta cruzada aprimora o compartilhamento de recursos entre contas, permitindo que você compartilhe um número ilimitado de tabelas com outra conta, eliminando os limites anteriores de associação de recursos por tipo de recurso. Para começar, atualize para a versão 5 de várias contas por meio do console ou da API do Lake Formation. Qualquer nova concessão de permissão entre contas usará automaticamente padrões curinga no compartilhamento de recursos em vez de associações individuais de recursos. Todos os compartilhamentos existentes entre contas continuam funcionando e todas as APIs existentes do Lake Formation permanecem compatíveis.

Benefícios da atualização de versões: a Cross-account v5 aprimora o compartilhamento entre contas, permitindo que você compartilhe centenas de milhares de tabelas entre contas.

Considerações ao atualizar as versões: novas concessões após a atualização da versão 5 adicionarão padrões de recursos curinga aos compartilhamentos de AWS recursos existentes do Resource Manager ou criarão novos compartilhamentos com padrões curinga. Depois de atualizado para a versão 5, o downgrade não é suportado.

Otimizar AWS RAM compartilhamentos de recursos

As novas versões (versão 2 e superior) de concessões entre contas utilizam de forma otimizada a AWS RAM capacidade para maximizar o uso de várias contas. Quando você compartilha um recurso com um diretor externo Conta da AWS ou do IAM, o Lake Formation pode criar um novo compartilhamento de recursos ou associar o recurso a um compartilhamento existente. Ao se associar aos compartilhamentos existentes, o Lake Formation reduz o número de convites de compartilhamento de recursos que um consumidor precisa aceitar. A versão 5 otimiza ainda mais o uso da RAM usando padrões de recursos baseados em curingas em vez de associações de recursos individuais, reduzindo significativamente as associações de recursos por compartilhamento de recursos.

Enable (Habilitar) AWS RAM ações via TBAC ou compartilhe recursos diretamente com os diretores

Para compartilhar recursos diretamente com entidades principais do IAM em outra conta, ou para habilitar compartilhamentos entre contas TBAC para organizações ou unidades organizacionais, é necessário atualizar as Configurações de versão entre contas para a versão 3. Para obter mais informações sobre limites AWS RAM de recursos, consulte Práticas recomendadas e considerações sobre compartilhamento de dados entre contas.

Permissões necessárias para atualizar a versão de Configurações Entre Contas

Se um concedente de permissão entre contas tiver permissões de política do IAM gerenciadas por `AWSLakeFormationCrossAccountManager`, não será necessária nenhuma configuração de permissão extra para o perfil de concedente ou entidade principal para permissão entre contas. No entanto, se o concedente entre contas não estiver usando a política gerenciada, o perfil do concedente ou entidade principal deverá ter as seguintes permissões do IAM concedidas para que a nova versão da concessão entre contas seja bem-sucedida.

JSON

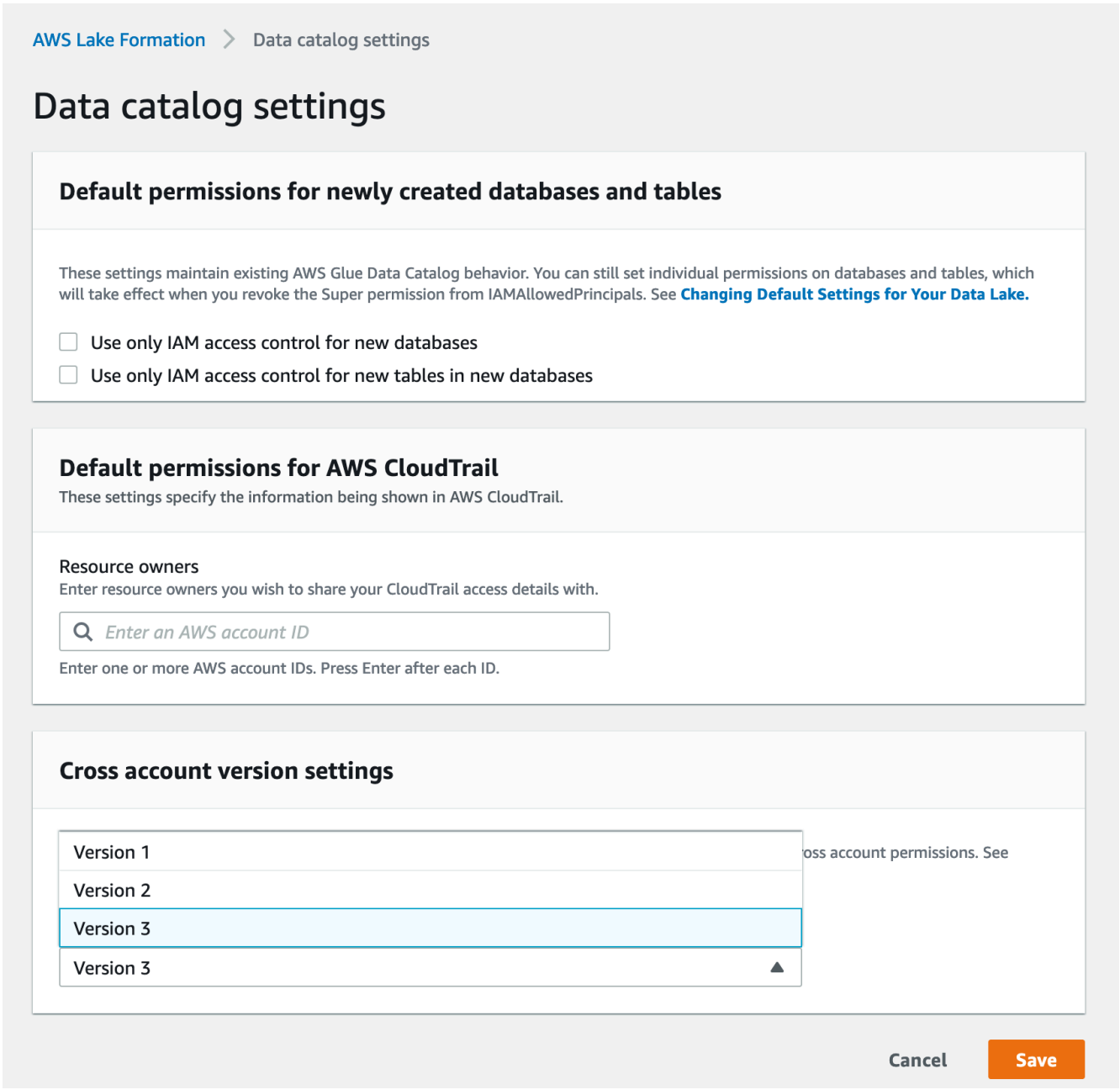
```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Sid": "VisualEditor1",
      "Effect": "Allow",
      "Action": [
         "ram:AssociateResourceShare",
         "ram:DisassociateResourceShare",
         "ram:GetResourceShares"
       ],
     "Resource": "*",
     "Condition": {
       "StringLike": {
         "ram:ResourceShareName": "LakeFormation*"
        }
       }
    }
  ]
}
```

Como habilitar a nova versão

Siga estas etapas para atualizar as configurações da versão da conta cruzada por meio do AWS Lake Formation console ou do AWS CLI.

Console

1. Escolha Versão 2, Versão 3, Versão 4 ou Versão 5 em Configurações de versão entre contas na página Configurações do catálogo de dados. Se você selecionar a Versão 1, o Lake Formation usará o modo padrão de compartilhamento de recursos.



2. Escolha Salvar.

AWS Command Line Interface (AWS CLI)

Use o `put-data-lake-settings` AWS CLI comando para definir o `CROSS_ACCOUNT_VERSION` parâmetro. Os valores aceitos são 1, 2, 3, 4 e 5.

```
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**
>
> Após escolher a Versão 2 ou a Versão 3, todas as novas concessões de recursos nomeados passarão pelo novo modo de concessão entre contas. Para otimizar o uso AWS RAM da capacidade de seus compartilhamentos entre contas existentes, recomendamos que você revogue as concessões que foram feitas com a versão mais antiga e conceda novamente no novo modo.

Compartilhando tabelas e bancos de dados do Catálogo de Dados entre Contas da AWS ou diretores do IAM de contas externas

Esta seção inclui instruções sobre como conceder permissões entre contas nos recursos do Catálogo de Dados a uma AWS conta externa, diretor do IAM, AWS organização ou unidade organizacional. A operação de concessão compartilha automaticamente esses recursos.

Tópicos

- Compartilhamento de dados usando controle de acesso baseado em tags
- Cross-account compartilhamento de dados usando o método de recurso nomeado

Compartilhamento de dados usando controle de acesso baseado em tags

AWS Lake Formation o controle de acesso baseado em tags (LF-TBAC) é uma estratégia de autorização que define permissões com base em atributos. As etapas a seguir explicam como conceder permissões entre contas usando LF-Tags.

Configuração obrigatória na producer/grantor conta

1. Adicionar LF-Tags.

   1. Faça login no console do Lake Formation como administrador ou LF-Tag criador do data lake.
   2. Na barra de navegação à esquerda, escolha Permissões e LF-Tags e permissões.
   3. Escolha Adicionar LF-Tag.

   Para obter instruções detalhadas sobre como criar LF-Tags, consulteCriando LF-Tags.

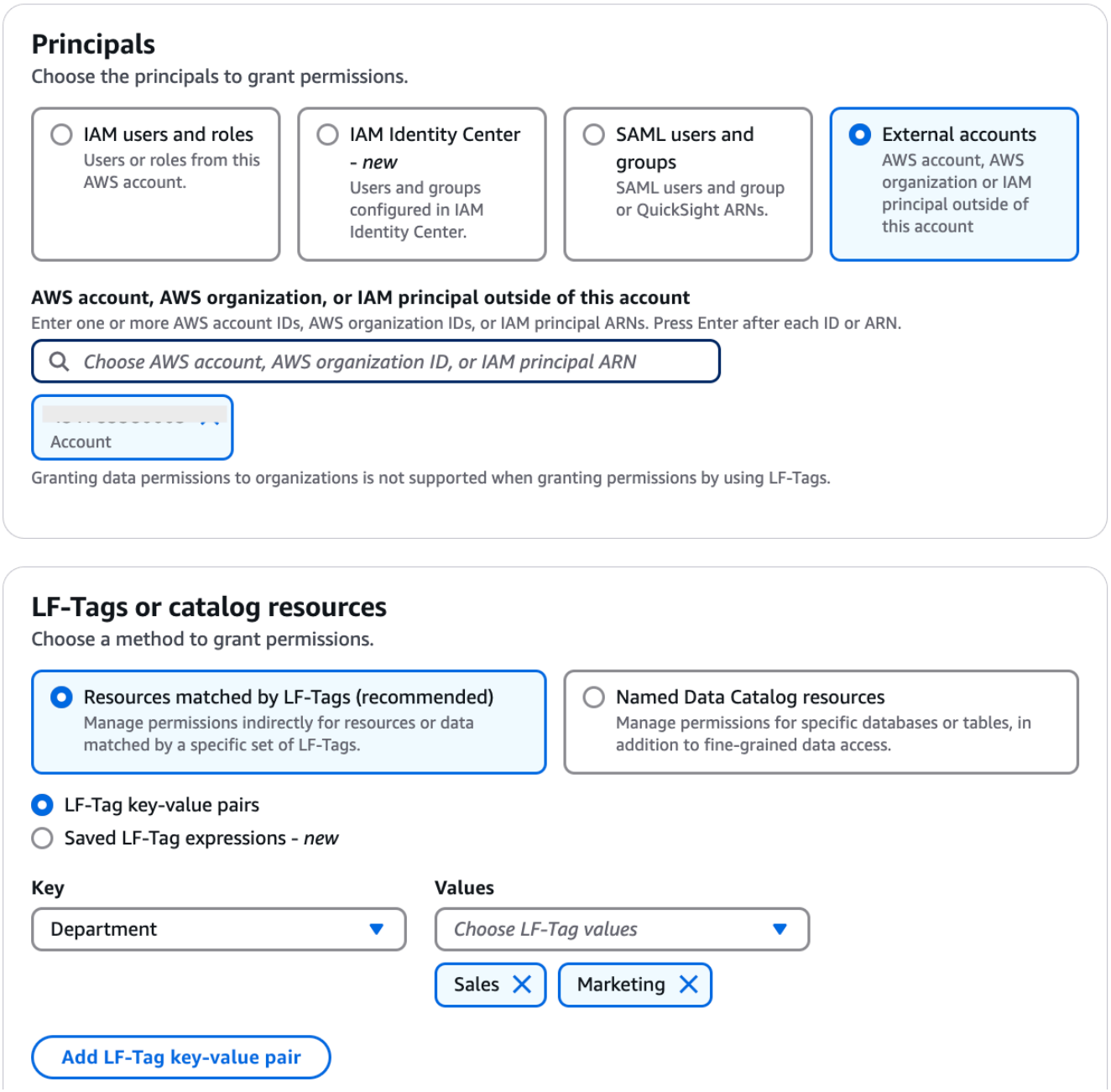
2. Conceda pares de LF-Tag valores-chave e chaves de permissões do Descreve and/or Associate aos diretores do IAM em sua conta ou em contas externas.

   A concessão de permissões em pares LF-Tag de valores-chave permite que os diretores visualizem e os atribuam aos recursos do LF-Tags Catálogo de Dados (bancos de dados, tabelas e colunas).

3. Em seguida, o administrador do data lake ou um diretor do IAM com permissão de associado pode atribuí-los LF-Tag a bancos de dados, tabelas ou colunas. Para obter mais informações, consulte Atribuição LF-Tags aos recursos do Catálogo de Dados.

4. Em seguida, conceda permissão de dados para contas externas usando LF-Tag expressões. Isso permite que o beneficiário ou o destinatário das permissões acesse os recursos do Data Catalog marcados com as mesmas chaves e valores.

   1. No painel de navegação, escolha Permissões e Permissões de dados.
   2. Selecione Conceder.
   3. Na página Conceder permissões, para Entidades principais, selecione Contas externas e insira o ID da Conta da AWS beneficiária, o perfil do IAM da entidade principal, ou o nome do recurso da Amazon (ARN) para a entidade principal (ARN da entidade principal), se estiver fazendo uma concessão direta entre contas a uma entidade principal externa. Você precisa pressionar Enter depois de inserir o ID da conta.



4. Para LF-Tags ou catalogue recursos, escolha Recursos correspondentes a LF-Tags (recomendado).

   a. Escolha a opção pares de LF-Tag valores-chave ou Expressões salvas LF-Tag .

   b. Se você escolher pares de LF-Tag chave-valor, insira a chave e o (s) valor (es) do LF-Tagque está associado ao recurso do Catálogo de Dados que está sendo compartilhado com a conta do beneficiário.

      O beneficiário recebe permissões nos recursos do Catálogo de Dados aos quais foi atribuída uma correspondência LF-Tag na LF-Tag expressão. Se a LF-Tag expressão

especificar vários valores por chave de tag, qualquer um dos valores de tag poderá corresponder.

5. Escolha as permissões em nível de banco de dados ou em nível de tabela a serem concedidas aos recursos que correspondam à expressão. LF-Tag

> **Important**
>
> Como o administrador do data lake deve conceder permissões em recursos compartilhados para as entidades principais na conta beneficiária, você sempre deve conceder permissões entre contas com a opção de concessão.

Para obter mais informações, consulte Conceder LF-Tag permissões usando o console.

> **Note**
>
> Entidades principais que receberem concessões diretas entre contas não terão a opção de Permissões concedíveis.

Configuração obrigatória na receiving/grantee conta

1. Faça login no console do Lake Formation como administrador do data lake da conta do consumidor.
2. Depois, receba o compartilhamento do recurso na conta do consumidor.
    1. Abra o AWS RAM console.
    2. No painel de navegação, selecione Compartilhados comigo, Compartilhamento de recursos.
    3. Selecione os compartilhamentos de recursos e escolha Aceitar compartilhamento de recursos.
3. Quando você compartilha um recurso com outra conta, o recurso ainda pertence à conta do produtor e não será visível no console do Athena. Para tornar o recurso visível no console do Athena, você precisa criar um link de recurso direcionando para o recurso compartilhado. Para obter instruções sobre como criar um link de recurso, consulte Como criar um link de recurso para uma tabela compartilhada do catálogo de dados e Como criar um link de recurso para um banco de dados compartilhado do catálogo de dados

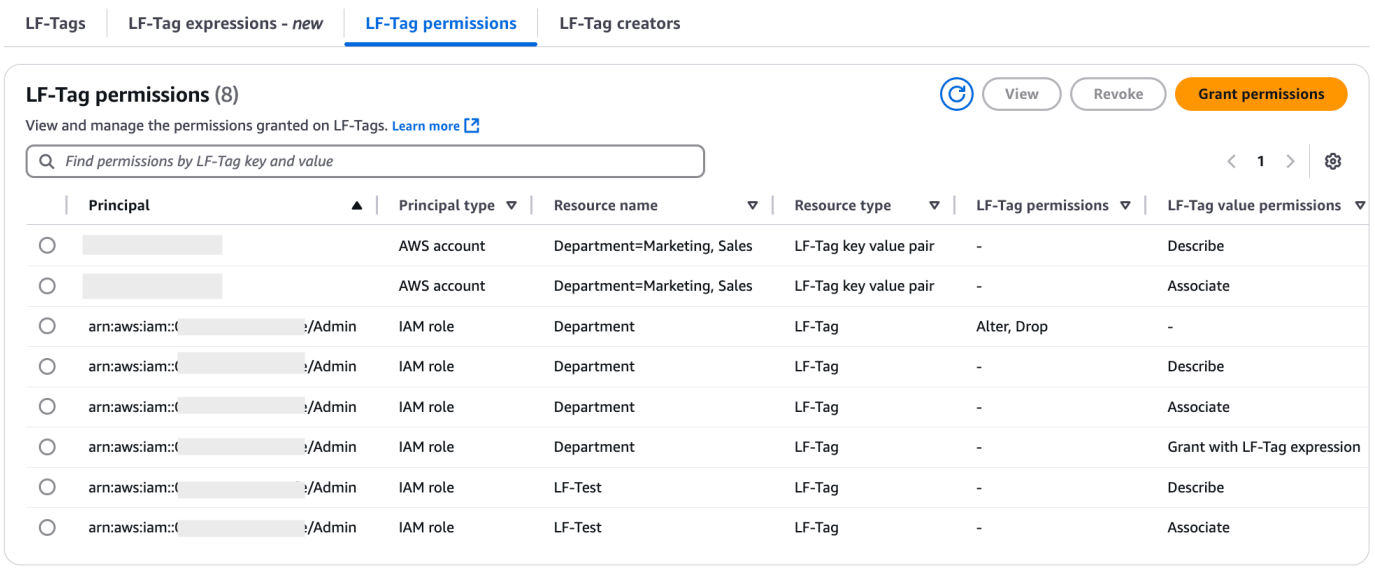
1. Escolha Bancos de dados ou Tabelas no Data Catalog.
2. Na Databases/Tables página, escolha Criar, link do recurso.
3. Insira as seguintes informações para um link do recurso de banco de dados:
   - Nome do link do recurso: um nome exclusivo para o link do recurso.
   - Catálogo de destino: o catálogo no qual você está criando o link do recurso.
   - Região do banco de dados compartilhado: a região do banco de dados compartilhado com você se você estiver criando o link de recurso em outra região.
   - Banco de dados compartilhado: escolha o banco de dados compartilhado.
   - ID do catálogo do banco de dados compartilhado: insira o ID do catálogo do banco de dados compartilhado.
4. Escolha Criar. É possível ver o link do recurso recém-criado na lista de bancos de dados.

Da mesma forma, você pode criar um link do recurso para a tabela compartilhada.

4. Agora, conceda a permissão Describe no link do recurso às entidades principais do IAM com as quais você está compartilhando o recurso.

   1. Na Databases/Tablespágina, selecione o link do recurso e, no menu Ações, escolha Conceder.
   2. Na seção Conceder permissões, selecione Usuários e perfis do IAM.
   3. Escolha o perfil do IAM ao qual você deseja conceder acesso ao link do recurso.
   4. Na seção de permissões Link do recurso, selecione Describe.
   5. Selecione Conceder.
5. Em seguida, conceda permissões LF-Tag de valor-chave aos diretores na conta do consumidor.

   Você deve conseguir encontrar os LF-Tags que são compartilhados com você na conta do consumidor no console do Lake Formation, em Permissões LF-Tags e permissões. É possível associar tags compartilhadas do concedente aos recursos compartilhados da conta do concedente que incluem: bancos de dados, tabelas e colunas. Você pode conceder permissões nos recursos a outras entidades principais.



1. No painel de navegação, em Permissões, Permissões de dados, selecione Conceder.
2. Na página Conceder permissões, selecione Usuários e perfis do IAM.
3. Em seguida, escolha os usuários e funções do IAM em sua conta para conceder acesso ao compartilhado databases/tables.
4. Em seguida, para obter LF-Tags ou catalogar recursos, escolha Recursos correspondentes a LF-Tags.
5. Em seguida, escolha a chave e os valores do LF-Tag que é compartilhado com você.
6. Depois, escolha as permissões de banco de dados e de tabela que você deseja conceder aos usuários e perfis do IAM. Você também pode escolher permissões concedíveis que permitem que os usuários e funções do IAM concedam permissões a outras pessoas. users/roles
7. Selecione Conceder.
8. É possível visualizar as concessões de permissão em Permissões de dados no console do Lake Formation.

Cross-account compartilhamento de dados usando o método de recurso nomeado

Você pode conceder permissões diretamente aos diretores em outra AWS conta, ou para uma conta externa Contas da AWS ou AWS Organizations. Conceder permissões do Lake Formation a organizações ou unidades organizacionais é equivalente a conceder a permissão a todos Conta da AWS nessa organização ou unidade organizacional.

Ao conceder permissões a contas ou organizações externas, você deve incluir a opção Permissões concedíveis. Somente o administrador do data lake na conta externa pode acessar os recursos compartilhados até que o administrador conceda permissões sobre os recursos compartilhados a outras entidades principais na conta externa.

> **Note**
>
> A opção de Permissões concedíveis não é suportada ao conceder permissões diretamente às entidades principais do IAM a partir de contas externas.

Siga as instruções em Conceder permissões de banco de dados usando o método de recurso nomeado para conceder permissões entre contas usando o método de recurso nomeado.

Conceder permissões em um banco de dados ou tabela compartilhada com sua conta

Depois que um recurso do Catálogo de Dados pertencente a outra AWS conta for compartilhado com sua AWS conta, como administrador do data lake, você poderá conceder permissões sobre o recurso compartilhado a outros diretores da sua conta. No entanto, você não pode conceder permissões sobre o recurso a outras contas AWS ou organizações.

Você pode usar o AWS Lake Formation console, a API ou o AWS Command Line Interface (AWS CLI) para conceder as permissões.

Como conceder permissões em um banco de dados compartilhado (método de recurso nomeado, console)

- Siga as instruções em Conceder permissões de banco de dados usando o método de recurso nomeado. Na lista Banco de dados em LF-Tags ou no catálogo de recursos, certifique-se de selecionar o banco de dados na conta externa, não um link de recurso para o banco de dados.

  Se você não encontrar o banco de dados na lista de bancos de dados, certifique-se de ter aceitado o convite de compartilhamento de recursos AWS Resource Access Manager (AWS RAM) para o banco de dados. Para obter mais informações, consulte Aceitando um convite de compartilhamento de recursos do AWS RAM.

Além disso, para obter as permissões `CREATE_TABLE` e `ALTER`, siga as instruções em Concessão de permissões de localização de dados (mesma conta) e não se esqueça de inserir o ID da conta proprietária no campo Localização da conta registrada.

Como conceder permissões em uma tabela compartilhada (método de recurso nomeado, console)

- Siga as instruções em Conceder permissões de tabela usando o método de recurso nomeado. Na lista Banco de dados em LF-Tags ou no catálogo de recursos, certifique-se de selecionar o banco de dados na conta externa, não um link de recurso para o banco de dados.

  Se você não encontrar tabela na lista de tabelas, certifique-se de ter aceitado o convite de compartilhamento de recursos AWS RAM para a tabela. Para obter mais informações, consulte Aceitando um convite de compartilhamento de recursos do AWS RAM.

  Além disso, para obter as permissões `ALTER`, siga as instruções em Concessão de permissões de localização de dados (mesma conta) e não se esqueça de inserir o ID da conta proprietária no campo Localização da conta registrada.

Para conceder permissões em recursos compartilhados (LF-TBAC método, console)

- Siga as instruções em Conceder permissões do catálogo de dados. Na seção LF-Tags ou recursos do catálogo, conceda a LF-Tag expressão exata que a conta externa concedeu à sua conta ou um subconjunto dessa expressão.

  Por exemplo, se uma conta externa concedeu a LF-Tag expressão `module=customers AND environment=production` à sua conta com a opção de concessão, como administrador do data lake, você pode conceder essa mesma expressão `module=customers` ou `environment=production` a um principal em sua conta. Você pode conceder somente as mesmas permissões ou um subconjunto das permissões do Lake Formation (por exemplo,`SELECT`,`ALTER`, etc.) que foram concedidas aos recursos por meio da LF-Tag expressão.

Para conceder permissões em uma tabela compartilhada (chamada método de recurso, AWS CLI)

- Digite um comando semelhante ao seguinte: Neste exemplo:

  - O ID AWS da sua conta é 1111-2222-3333.

- A conta que possui a tabela e que a concedeu à sua conta é 1234-5678-9012.
- A permissão `SELECT` está sendo concedida na tabela compartilhada `pageviews` ao usuário `datalake_user1`. Esse usuário é a entidade principal da sua conta.
- A tabela `pageviews` está no banco de dados `analytics`, que pertence à conta 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"}}'
```

Observe que a conta proprietária deve ser especificada na propriedade `CatalogId`, no argumento`resource`.

Como conceder permissões de links de recursos

Siga estas etapas para conceder AWS Lake Formation permissões em um ou mais links de recursos a um diretor em sua AWS conta.

Após criar um link de recurso, somente você poderá visualizá-lo e acessá-lo. (Isso pressupõe que Usar somente o controle de acesso do IAM para novas tabelas nesse banco de dados não esteja habilitado para o banco de dados.) Para permitir que outras entidades principais da sua conta acessem o link do recurso, conceda pelo menos a permissão `DESCRIBE`.

> **Important**
>
> Conceder permissões em um link de recurso não concede permissões no banco de dados ou tabela de destino (vinculado). Você deve conceder permissões no destino separadamente.

Você pode conceder permissões usando o console do Lake Formation, a API ou o AWS Command Line Interface (AWS CLI).

console

Como conceder permissões de links de recursos usando o console do Lake Formation

1. Execute um destes procedimentos:

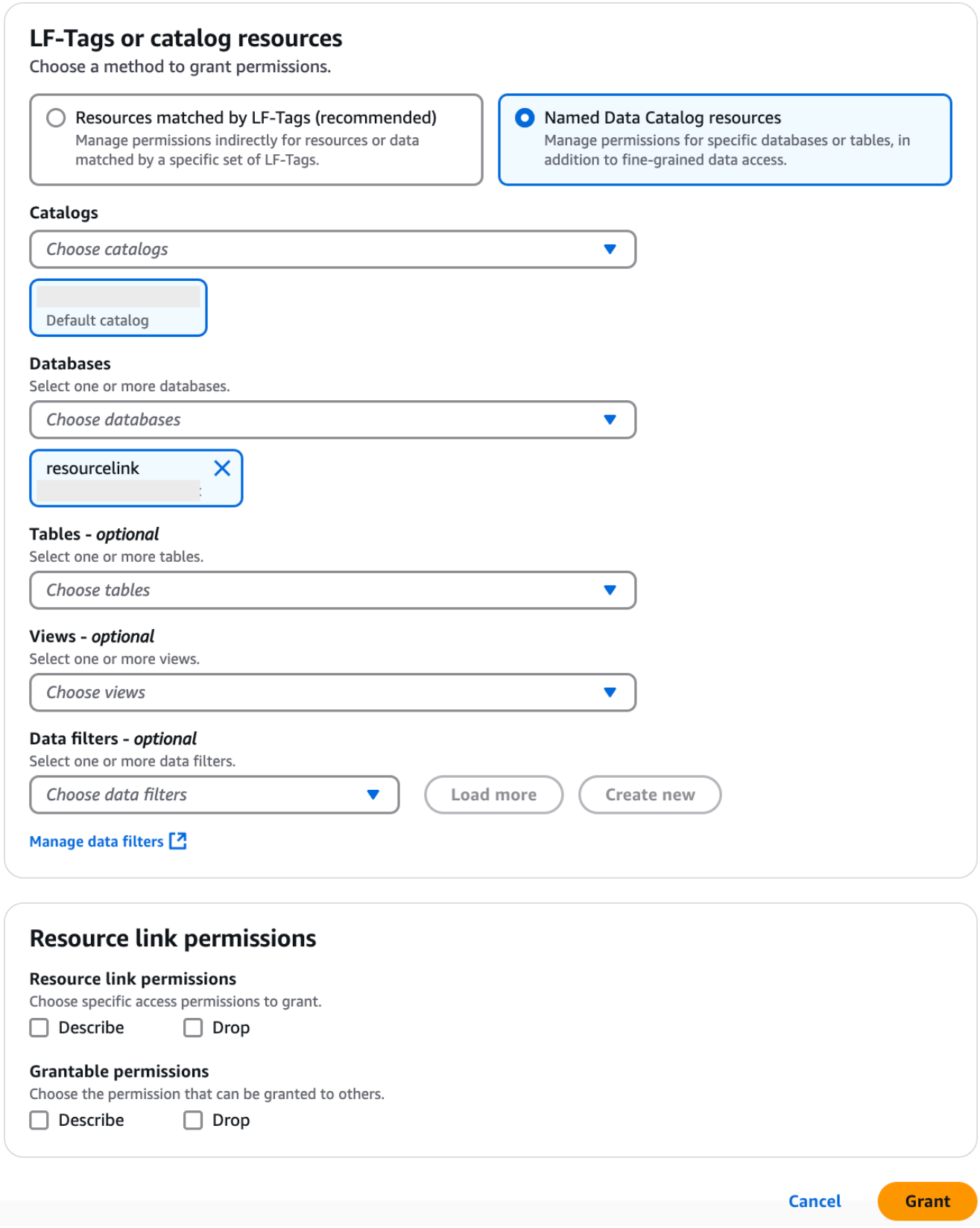
   - Para obter links de recursos de banco de dados, siga as etapas em Conceder permissões de banco de dados usando o método de recurso nomeado para fazer o seguinte:

     1. Selecione o link do recurso na lista de bancos de dados no Data Catalog, Bancos de dados.
     2. Escolha Conceder para abrir a página Conceder permissões.
     3. Especifique as entidades principais para conceder permissões.
     4. Os campos Catálogos e Bancos de Dados são preenchidos.

   - Para obter links de recursos de tabela, siga as etapas em Conceder permissões de tabela usando o método de recurso nomeado para fazer o seguinte:

     1. Selecione o link do recurso na lista de tabelas no Data Catalog, Tabelas.
     2. Abra a página Conceder permissões.
     3. Especifique as entidades principais.
     4. Os campos Catálogos, Bancos de Dados e Tabelas são preenchidos.
     5. Especifique as entidades principais.

2. Em Permissões, selecione as permissões a serem concedidas. Como opção, selecione permissões concedíveis.



3. Selecione Conceder.

AWS CLI

Para conceder permissões de links de recursos usando AWS CLI

- Execute o comando `grant-permissions`, especificando um link de recurso como recurso.

  Example

  Este exemplo concede `DESCRIBE` ao usuário `datalake_user1` na tabela o link do recurso no banco de dados `incidents-link` `issues` na AWS conta 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"}}'
```

Consulte também:

- Criação de links de recursos
- Referência de permissões do Lake Formation

Como acessar os dados subjacentes de uma tabela compartilhada

Suponha que a AWS conta A compartilhe uma tabela do Catálogo de Dados com a conta B — por exemplo, concedendo a opção de concessão `SELECT` na tabela à conta B. Para que um principal na conta B possa ler os dados subjacentes da tabela compartilhada, as seguintes condições devem ser atendidas:

- O administrador do data lake na conta B deve aceitar o compartilhamento. (Isso não é necessário se as contas A e B estiverem na mesma organização ou se a concessão tiver sido feita com o método de controle de acesso baseado em tags do Lake Formation.)
- O administrador do data lake deve conceder novamente à entidade principal a permissão `SELECT` do Lake Formation que a conta A concedeu na tabela compartilhada.
- A entidade principal deve ter as seguintes permissões de IAM na tabela, no banco de dados que a contém e no catálogo de dados da conta A.

> **Note**
>
> Na seguinte política do IAM:
>
> - *<account-id-A>*Substitua pelo AWS ID da conta A.
> - *<region>*Substitua por uma região válida.
> - *<database>*Substitua pelo nome do banco de dados na conta A que contém a tabela compartilhada.
> - *<table>*Substitua pelo nome da tabela compartilhada.

JSON

```
{
    "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:us-east-1:111122223333:table/<database>/<table>",
            "arn:aws:glue:us-east-1:111122223333:database/<database>",
            "arn:aws:glue:us-east-1:111122223333:catalog"
           ]
        },
        {
          "Effect": "Allow",
          "Action": [
            "lakeformation:GetDataAccess"
           ],
          "Resource": [
            "*"
```

```
            ]
        }
    ]
}
```

Consulte também:

- Aceitando um convite de compartilhamento de recursos do AWS RAM

Cross-account CloudTrail registro

O Lake Formation fornece uma trilha de auditoria centralizada de todo o acesso entre contas aos dados em seu data lake. Quando uma AWS conta de destinatário acessa dados em uma tabela compartilhada, o Lake Formation copia o CloudTrail evento para os registros da CloudTrail conta proprietária. Os eventos copiados incluem consultas aos dados por serviços integrados, como o Amazon Redshift Spectrum, Amazon Athena e acessos a dados por trabalhos. AWS Glue

CloudTrail os eventos para operações entre contas nos recursos do Catálogo de Dados são copiados de forma semelhante.

Como proprietário do recurso, se você habilitar o registro em nível de objeto no Amazon S3, poderá executar consultas que unem eventos do S3 aos CloudTrail eventos do CloudTrail Lake Formation para determinar as contas que acessaram seus buckets do S3.

Tópicos

- Incluindo identidades principais em registros de várias contas CloudTrail
- Consultando CloudTrail registros para acesso entre contas do Amazon S3

Incluindo identidades principais em registros de várias contas CloudTrail

Por padrão, CloudTrail os eventos entre contas adicionados aos registros do destinatário do recurso compartilhado e copiados para os registros do proprietário do recurso contêm somente o ID AWS principal do responsável externo da conta, não o nome de recurso da Amazon (ARN) legível por humanos do principal (ARN principal). Ao compartilhar recursos dentro de limites confiáveis, como dentro da mesma organização ou equipe, você pode optar por incluir o ARN principal nos CloudTrail

eventos. As contas do proprietário do recurso podem então rastrear as entidades principais nas contas de destinatários que acessam seus recursos próprios.

> **Important**
>
> Como destinatário do recurso compartilhado, para ver o ARN principal em eventos em seus próprios CloudTrail registros, você deve optar por compartilhar o ARN principal com a conta do proprietário.
> Se o acesso aos dados ocorrer por meio de um link de recurso, dois eventos serão registrados na conta do destinatário do recurso compartilhado: um para o acesso ao link do recurso e outro para o acesso ao recurso de destino. O evento para o acesso ao link de recurso inclui o ARN da entidade principal. O evento para o acesso ao recurso de destino não inclui o ARN da entidade principal sem selecionar a opção de inclusão. O evento de acesso ao link do recurso não é copiado para a conta do proprietário.

A seguir está um trecho de um CloudTrail evento padrão entre contas (sem aceitação). A conta que executa o acesso aos dados é 1111-2222-3333. Este é o log mostrado na conta de origem e na conta do proprietário do recurso. Lake Formation preenche logs em ambas as contas no caso entre contas.

```
{
    "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"
    },
...
}
```

Como consumidor de recursos compartilhados, quando você opta por incluir o ARN da entidade principal, o trecho se torna o seguinte. O campo `lakeFormationPrincipal` representa o perfil final ou o usuário que executa a consulta por meio do Amazon Athena, Amazon Redshift Spectrum ou trabalhos do 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"
    },
...
}
```

Para optar por incluir ARNs principais nos registros de várias contas CloudTrail

1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

   Faça login como usuário `Administrator` ou como usuário com a política do IAM `Administrator Access`.
2. No painel de navegação, selecione Configurações.
3. Na página Configurações do catálogo de dados, na AWS CloudTrail seção Permissões padrão para, para proprietários de recursos, insira uma ou mais IDs de conta do proprietário do AWS recurso.

   Pressione Enter após cada ID da conta.
4. Escolha Salvar.

   Agora, CloudTrail os eventos entre contas armazenados nos registros do destinatário do recurso compartilhado e do proprietário do recurso contêm o ARN principal.

Consultando CloudTrail registros para acesso entre contas do Amazon S3

Como proprietário de um recurso compartilhado, você pode consultar CloudTrail os registros do S3 para determinar as contas que acessaram seus buckets do Amazon S3 (desde que você tenha habilitado o registro em nível de objeto no Amazon S3). Isso se aplica somente aos locais do S3 que você registrou no Lake Formation. Se os consumidores de recursos compartilhados optarem por incluir os ARNs principais nos CloudTrail registros do Lake Formation, você poderá determinar as funções ou os usuários que acessaram os buckets.

Ao executar consultas com Amazon Athena, você pode unir eventos do Lake Formation e CloudTrail eventos do S3 na CloudTrail propriedade do nome da sessão. As consultas também podem filtrar eventos do Lake Formation em `eventName="GetDataAccess"` e eventos do S3 em `eventName="Get Object"` ou `eventName="Put Object"`.

A seguir está um trecho de um CloudTrail evento entre contas do Lake Formation em que dados em um local registrado do S3 foram acessados.

```
{
  "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"
   }
}
```

O valor da `lakeFormationRoleSessionName` chave`AWSLF-00-GL-111122223333-B8JSAjo5QA`,, pode ser associado ao nome da sessão na `principalId` chave do CloudTrail evento do S3. A seguir está um trecho do evento CloudTrail S3. Veja a localização do nome da sessão.

```
{
   "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"
     },
   ..............
   ..............
}
```

O nome da sessão é formatado da seguinte forma:

```
AWSLF-<version-number>-<query-engine-code>-<account-id->-<suffix>
```

`version-number`

A versão desse formato, atualmente `00`. Se o formato do nome da sessão mudar, a próxima versão será `01`.

`query-engine-code`

Indica a entidade que acessou os dados. Os valores atuais são:

`GL`	Tarefa de ETL AWS Glue
`AT`	Athena
`RE`	Amazon Redshift Spectrum

`account-id`

O ID da AWS conta que solicitou as credenciais do Lake Formation.

`suffix`

Uma string gerada aleatoriamente.

Gerenciando permissões entre contas usando ambos AWS Glue e Lake Formation

É possível conceder acesso entre contas a recursos do catálogo de dados e dados subjacentes usando o AWS Glue ou o AWS Lake Formation.

No AWS Glue, você concede permissões entre contas criando ou atualizando uma política de recursos do Catálogo de Dados. No Lake Formation, você concede permissões entre contas usando o modelo de permissões `GRANT/REVOKE` do Lake Formation e a operação da API `Grant Permissions`.

> **Tip**
>
> Recomendamos que você confie somente nas permissões do Lake Formation para proteger seu data lake.

Você pode ver os subsídios entre contas do Lake Formation usando o console do Lake Formation ou o console AWS Resource Access Manager (AWS RAM). No entanto, essas páginas do console não mostram permissões entre contas concedidas pela política de recursos do catálogo de dados do AWS Glue. Da mesma forma, você pode visualizar as concessões entre contas na política de recursos do catálogo de dados usando a página Configurações do console AWS Glue, mas essa página não mostra as permissões entre contas concedidas usando o Lake Formation.

Para garantir que você não perca nenhuma concessão ao visualizar e gerenciar permissões entre contas, o Lake Formation e AWS Glue exigem que você execute as seguintes ações para indicar que está ciente e permite concessões entre contas pelo Lake Formation e AWS Glue.

Ao conceder permissões entre contas usando o AWS Glue Política de recursos do Catálogo de Dados

Se sua conta (conta de concedente ou conta de produtor) não tiver feito concessões entre contas que usam o AWS RAM para compartilhar os recursos, você pode salvar uma política de recursos do Catálogo de Dados normalmente no AWS Glue. No entanto, se concessões que envolvem compartilhamentos de AWS RAM recursos já tiverem sido feitas, você deverá fazer o seguinte para garantir que o salvamento da política de recursos seja bem-sucedido:

- Quando você salva a política de recursos na página Configurações do console do AWS Glue, o console emite um alerta informando que as permissões na política serão adicionais às permissões

concedidas usando o console do Lake Formation. Você deve escolher Continuar para salvar a política.

- Ao salvar a política de recursos usando a operação da API `glue:PutResourcePolicy`, você deve definir o campo `EnableHybrid` como 'TRUE' (type = string).

  Para atualizar uma política de recursos existente, use a operação de API `glue:GetResourcePolicy` para recuperar sua política atual primeiro e depois modificá-la conforme necessário antes de chamar `glue:PutResourcePolicy`.

  > **Note**
  >
  > Ao criar políticas AWS Glue de recursos para acesso entre contas, conceda somente as permissões mínimas necessárias para seu caso de uso específico.

  Para obter mais informações, consulte PutResourcePolicy Action (Python: put_resource_policy) no Guia do desenvolvedor.AWS Glue

Ao conceder permissões entre contas usando o método de recursos nomeados do Lake Formation

Se não houver uma política de recursos do Catálogo de Dados em sua conta (conta de produtor), as concessões entre contas do Lake Formation que você fizer procederão normalmente. No entanto, se existir uma política de recursos do catálogo de dados, você deverá adicionar a seguinte declaração a ela para permitir que suas concessões entre contas sejam bem-sucedidas se forem feitas com o método de recurso nomeado. *<region>*Substitua por um nome de região válido e *<account-id>* pelo ID AWS da sua conta (ID da conta do produtor).

```
{
  "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"
```

```
    ]
}
```

Sem essa declaração adicional, o subsídio do Lake Formation é bem-sucedido, mas fica bloqueado AWS RAM e a conta do destinatário não pode acessar o recurso concedido.

> **Important**
>
> Ao usar o método de controle de acesso (LF-TBAC) baseado em tags do Lake Formation para fazer concessões entre contas, você deve ter uma política de recursos do Catálogo de Dados com pelo menos as permissões especificadas em. Pré-requisitos

> Consulte também:
>
> - Controle de acesso a metadados(para uma discussão sobre o método de recurso nomeado versus o método de controle de acesso (LF-TBAC) baseado em tags Lake Formation).
> - Visualizando tabelas e bancos de dados compartilhados do catálogo de dados
> - Como trabalhar com as configurações do catálogo de dados no console AWS Glue no Guia do Desenvolvedor AWS Glue
> - Concedendo Cross-Account acesso no Guia do AWS Glue desenvolvedor (para exemplos de políticas de recursos do Catálogo de Dados)

Visualizando todas as concessões entre contas usando a operação de GetResourceShares API

Se sua empresa concede permissões entre contas usando uma política de AWS Glue Data Catalog recursos e subsídios do Lake Formation, a única maneira de visualizar todas as concessões entre contas em um só lugar é usando a operação de `glue:GetResourceShares` API.

Quando você concede permissões do Lake Formation em todas as contas usando o método de recurso nomeado, AWS Resource Access Manager (AWS RAM) cria uma política de recursos AWS Identity and Access Management (IAM) e a armazena em sua AWS conta. A política concede as permissões necessárias para acessar o recurso. AWS RAM cria uma política de recursos separada

para cada concessão entre contas. Você pode ver todas essas políticas usando a operação da API `glue:GetResourceShares`.

> **Note**
>
> Essa operação também restaura a política de recursos do catálogo de dados. No entanto, se você habilitou a criptografia de metadados nas configurações do Catálogo de Dados e não tem permissão na AWS KMS chave, a operação não retornará a política de recursos do Catálogo de Dados.

Como visualizar todas as concessões entre contas

- Digite o AWS CLI comando a seguir.

```
aws glue get-resource-policies
```

A seguir está um exemplo de política de recursos que AWS RAM cria e armazena quando você concede permissões na tabela `t` no banco de dados `db1` para a AWS conta 1111-2222-3333.

JSON

```
{
  "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:us-east-1:111122223333:table/db1/t"
     ]
    }
  ]
}
```

Consulte também:

- GetResourceShares Ação (Python: get_resource_policies) no Guia do desenvolvedor AWS Glue

Acessar e visualizar tabelas e bancos de dados compartilhados do catálogo de dados

Para o administrador do data lake e para os diretores que receberam permissões, os recursos que são compartilhados com sua AWS conta aparecem no Catálogo de Dados como se fossem recursos em sua conta. O console exibe a conta que tem o recurso.

Você pode visualizar os recursos que são compartilhados com sua conta usando o console do Lake Formation. Você também pode usar o console AWS Resource Access Manager (AWS RAM) para visualizar os recursos que são compartilhados com sua conta e os recursos que você compartilhou com outras AWS contas usando o método de recurso nomeado.

Important

Quando alguém usa o método de recurso nomeado para conceder permissões entre contas em um recurso do Catálogo de Dados para sua conta ou AWS organização, o Lake Formation usa o serviço AWS Resource Access Manager (AWS RAM) para compartilhar o recurso. Se sua conta estiver na mesma AWS organização da conta concedente, o recurso compartilhado estará disponível para você imediatamente.
No entanto, se sua conta não estiver na mesma organização, AWS RAM envia um convite à sua conta para aceitar ou rejeitar o compartilhamento de recursos. Em seguida, para disponibilizar o recurso compartilhado, o administrador do data lake em sua conta deve usar o AWS RAM console ou a CLI para aceitar o convite.

O console do Lake Formation exibe um alerta se houver um convite de compartilhamento de AWS RAM recursos aguardando para ser aceito. Somente usuários autorizados a ver os AWS RAM convites recebem o alerta.

Consulte também:

- Compartilhando tabelas e bancos de dados do Catálogo de Dados entre AWS Contas
- Cross-account compartilhamento de dados em Lake Formation
- Como acessar os dados subjacentes de uma tabela compartilhada
- Controle de acesso a metadados(para obter informações sobre o método de recurso nomeado versus o LF-TBAC método para compartilhar recursos.)

Tópicos

- Aceitando um convite de compartilhamento de recursos do AWS RAM
- Visualizando tabelas e bancos de dados compartilhados do catálogo de dados

Aceitando um convite de compartilhamento de recursos do AWS RAM

Se um recurso do Catálogo de Dados for compartilhado com sua AWS conta e sua conta não estiver na mesma AWS organização da conta de compartilhamento, você não terá acesso ao recurso compartilhado até aceitar um convite de compartilhamento de recursos de AWS Resource Access Manager (AWS RAM). Como administrador do data lake, você deve primeiro consultar AWS RAM os convites pendentes e depois aceitar o convite.

Você pode usar o AWS RAM console, a API ou AWS Command Line Interface (AWS CLI) para ver e aceitar convites.

Para ver e aceitar um convite de compartilhamento de recursos do AWS RAM (console)

1. Certifique-se de ter as permissões AWS Identity and Access Management (IAM) necessárias para visualizar e aceitar convites de compartilhamento de recursos.

   Para obter informações sobre as políticas de IAM sugeridas para administradores de data lake, consulte the section called “Permissões de administrador do data lake”.

2. Siga as instruções em Aceitar e rejeitar convites no Guia do usuário do AWS RAM .

Para ver e aceitar um convite de compartilhamento de recursos do AWS RAM (AWS CLI)

1. Certifique-se de ter as permissões AWS Identity and Access Management (IAM) necessárias para visualizar e aceitar convites de compartilhamento de recursos.

   Para obter informações sobre as políticas de IAM sugeridas para administradores de data lake, consulte the section called “Permissões de administrador do data lake”.

2. Insira o comando a seguir para visualizar os convites de compartilhamento de recursos pendentes.

   ```
   aws ram get-resource-share-invitations
   ```

   A saída deve ser semelhante a esta.

   ```
   {
       "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"
           }
       ]
   }
   ```

   Observe o status de `PENDING`.

3. Copie o valor da chave `resourceShareInvitationArn` para a área de transferência.

4. Cole o valor no comando a seguir*<invitation-arn>*, substitua-o e insira o comando.

   ```
   aws ram accept-resource-share-invitation --resource-share-invitation-
   arn <invitation-arn>
   ```

A saída deve ser semelhante a esta.

```
{
    "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"
        }
    ]
}
```

Observe o status de `ACCEPTED`.

Visualizando tabelas e bancos de dados compartilhados do catálogo de dados

Você pode visualizar os recursos que são compartilhados com sua conta usando o console do Lake Formation ou a AWS CLI. Você também pode usar o console AWS Resource Access Manager (AWS RAM) ou a CLI para visualizar os recursos que são compartilhados com sua conta e os recursos que você compartilhou com outras AWS contas.

Para visualizar recursos compartilhados usando o console do Lake Formation

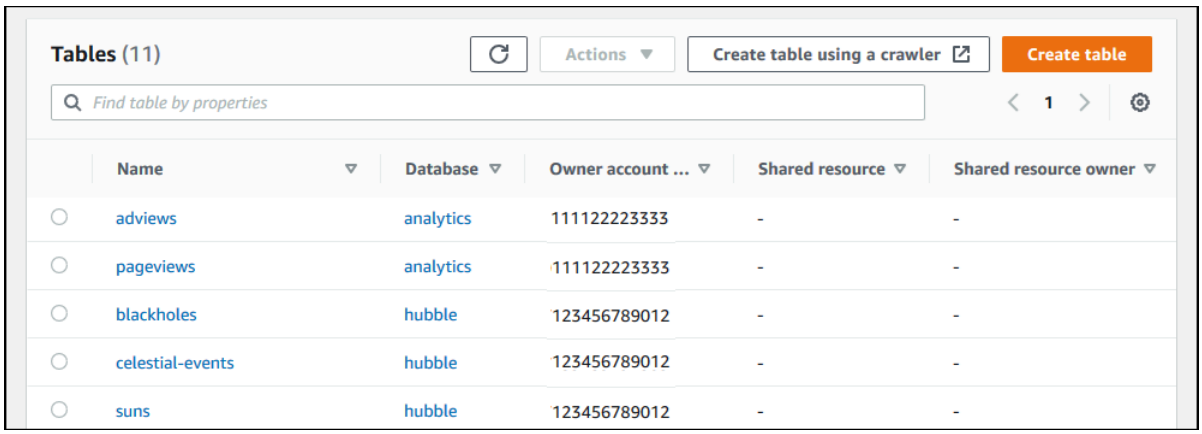
1. Abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.

   Faça login como administrador do data lake ou como usuário que recebeu permissões em uma tabela compartilhada.
2. Para visualizar os recursos que são compartilhados com sua AWS conta, faça o seguinte:

   - Para visualizar tabelas que são compartilhadas com sua conta, escolha Tabelas no painel de navegação.

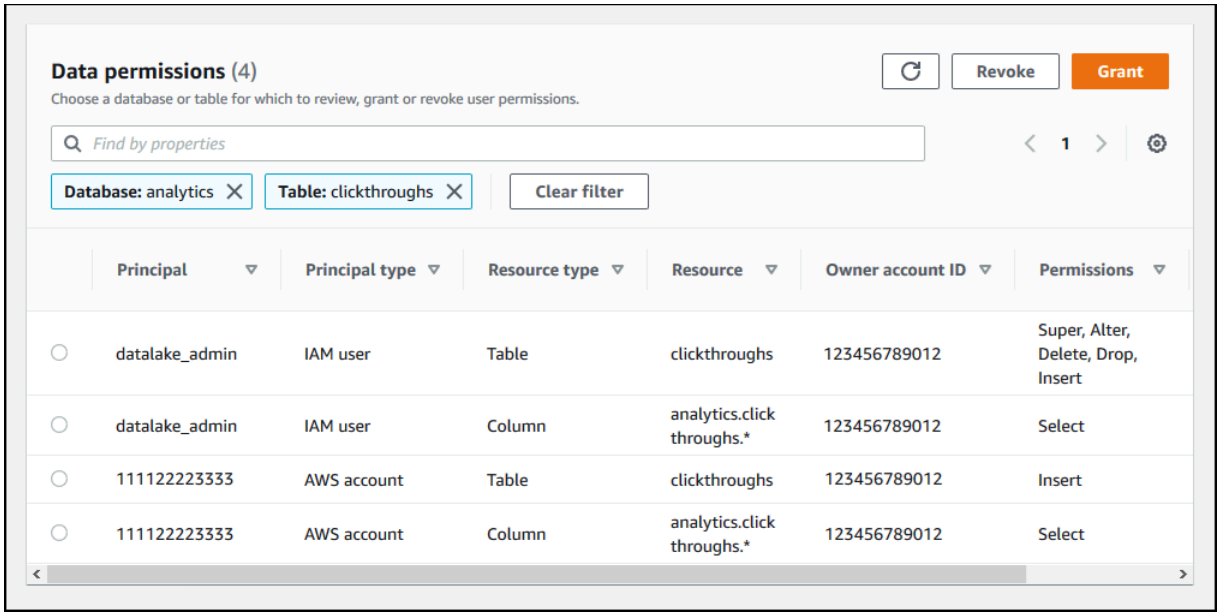
- Para visualizar tabelas que são compartilhadas com sua conta, escolha Banco de dados no painel de navegação.

O console exibe uma lista de bancos de dados ou tabelas em sua conta e compartilhados com sua conta. Para recursos compartilhados com sua conta, o console exibe o ID da conta AWS do proprietário na coluna ID da conta do proprietário (a terceira coluna na captura de tela a seguir).



3. Para visualizar os recursos que você compartilhou com outras AWS contas ou organizações, no painel de navegação, escolha Permissões de dados.

Os recursos que você compartilhou estão listados na página Permissões de dados com o número da conta externa mostrado na coluna Entidade principal, conforme mostrado na imagem a seguir.



Para visualizar recursos compartilhados usando o AWS RAM console

1. Certifique-se de ter as permissões AWS Identity and Access Management (IAM) necessárias para visualizar os recursos compartilhados usando AWS RAM.

   No mínimo, você deve ter a permissão `ram:ListResources`. Essa permissão está incluída na política gerenciada pela AWS `AWSLakeFormationCrossAccountManager`.
2. Faça login no Console de gerenciamento da AWS e abra o AWS RAM console em https://console.aws.amazon.com/ram/home.
3. Execute um destes procedimentos:

   - Para ver os recursos que você compartilhou, no painel de navegação, em Compartilhado por mim, escolha Recursos compartilhados.
   - Para ver os recursos compartilhados com você, no painel de navegação, em Compartilhado comigo, escolha Recursos compartilhados.

Criação de links de recursos

Links de recursos são objetos do Catálogo de Dados que são links para bancos de dados e tabelas de metadados — normalmente para bancos de dados e tabelas compartilhados de outras contas. AWS Eles ajudam a permitir o acesso entre contas aos dados no data lake em todas as AWS regiões.

> **Note**
>
> O Lake Formation suporta a consulta de tabelas do Catálogo de Dados em todas as AWS regiões. Você pode acessar os bancos de dados e tabelas do Catálogo de Dados de qualquer AWS região criando links de recursos nessas regiões que apontam para bancos de dados e tabelas compartilhados em diferentes regiões.

Tópicos

- Como os links de recursos funcionam no Lake Formation
- Como criar um link de recurso para uma tabela compartilhada do catálogo de dados
- Como criar um link de recurso para um banco de dados compartilhado do catálogo de dados
- Manipulação de links de recursos em AWS Glue APIs

Como os links de recursos funcionam no Lake Formation

Um link de recurso é um objeto do catálogo de dados que é um link para um banco de dados, ou uma tabela local, ou compartilhada. Depois de criar um link de recurso para um banco de dados ou tabela, você pode usar o nome do link de recurso onde quer que use o nome do banco de dados ou da tabela. Juntamente com as tabelas que você possui ou que são compartilhadas com você, os links de recursos de tabela são retornados pelo `glue:GetTables()` e aparecem como entradas na página Tabelas do console do Lake Formation. Os links de recursos para bancos de dados agem de maneira semelhante.

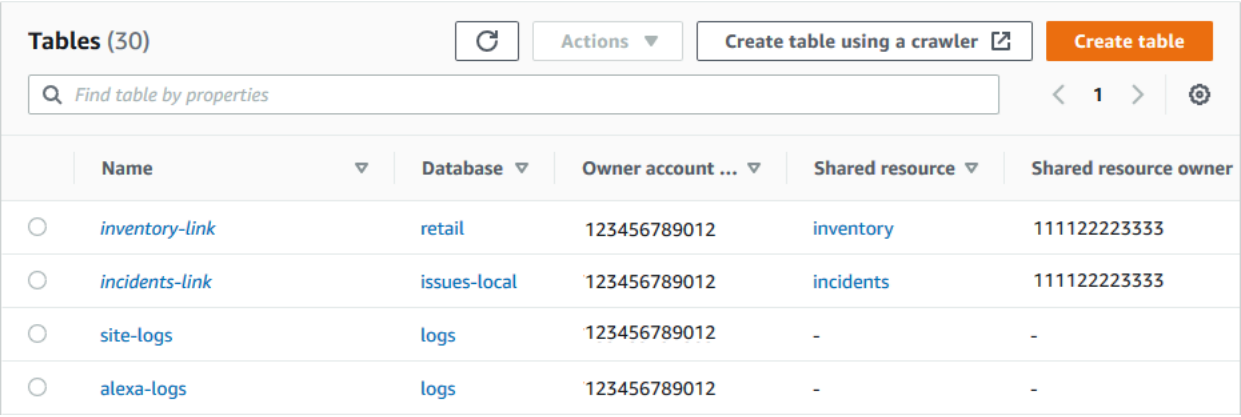
A criação de um link de recurso para um banco de dados ou tabela permite que você:

- Atribua um nome diferente a um banco de dados ou tabela em seu catálogo de dados. Isso é especialmente útil se AWS contas diferentes compartilharem bancos de dados ou tabelas com o mesmo nome, ou se vários bancos de dados em sua conta tiverem tabelas com o mesmo nome.
- Acesse os bancos de dados e tabelas do Catálogo de Dados de qualquer AWS região criando links de recursos nessas regiões apontando para o banco de dados e tabelas em outra região. Você pode executar consultas em qualquer região com esses links de recursos usando o Athena, o Amazon EMR e executar trabalhos do AWS Glue no ETL Spark, sem copiar os dados de origem nem os metadados no catálogo de dados do Glue.
- Use AWS serviços integrados, como Amazon Athena o Amazon Redshift Spectrum, para executar consultas que acessam bancos de dados ou tabelas compartilhados. Alguns serviços integrados não podem acessar diretamente bancos de dados ou tabelas entre contas. No entanto, eles podem acessar links de recursos em sua conta para bancos de dados e tabelas em outras contas.

> **Note**
>
> Você não precisa criar um link de recurso para referenciar um banco de dados ou tabela compartilhados em scripts de extração, transformação e carregamento (ETL) do AWS Glue. No entanto, para evitar ambiguidades quando várias contas da AWS compartilham um banco de dados ou tabela com o mesmo nome, você pode criar e usar um link de recurso ou especificar a ID do catálogo ao invocar operações de ETL.

O exemplo a seguir mostra a página Tabelas do console Lake Formation, que lista dois links de recursos. Os nomes dos links de recursos são sempre exibidos em itálico. Cada link de recurso é exibido junto com o nome e o proprietário do recurso compartilhado vinculado. Neste exemplo, um

administrador de data lake na AWS conta 1111-2222-3333 compartilhou as tabelas `inventory` e `incidents` com a conta 1234-5678-9012. Em seguida, um usuário dessa conta criou links de recursos para essas tabelas compartilhadas.



A seguir estão notas e restrições sobre links de recursos:

- Os links de recursos são necessários para permitir que serviços integrados, como Athena e Redshift Spectrum, consultem os dados subjacentes das tabelas compartilhadas. As consultas nesses serviços integrados são construídas com base nos nomes dos links de recursos.
- Supondo que a configuração Usar somente o controle de acesso do IAM para novas tabelas nesse banco de dados esteja desativada para o banco de dados que o contém, somente a entidade principal que criou um link de recurso pode visualizá-lo e acessá-lo. Para permitir que outras entidades principais da sua conta acessem um link de recurso, conceda a permissão `DESCRIBE` nela. Para permitir que outras pessoas descartem um link de recurso, conceda a permissão `DROP` para ele. Os administradores do data lake podem acessar todos os links de recursos na conta. Para descartar um link de recurso criado por outra entidade principal, o administrador do data lake deve primeiro conceder a si a permissão `DROP` no link do recurso. Para obter mais informações, consulte Referência de permissões do Lake Formation.

  > **Important**
  >
  > Conceder permissões em um link de recurso não concede permissões no banco de dados ou tabela de destino (vinculado). Você deve conceder permissões no destino separadamente.

- Para criar um link de recurso, você precisa da `CREATE_DATABASE` permissão `CREATE_TABLE` ou do Lake Formation, bem como da permissão `glue:CreateTable` or `glue:CreateDatabase` AWS Identity and Access Management (IAM).

- Você pode criar links de recursos para recursos locais (próprios) do Catálogo de Dados, bem como para recursos compartilhados com sua AWS conta.
- Quando você cria um link de recurso, nenhuma verificação é executada para ver se o recurso compartilhado de destino existe ou se você tem permissões entre contas no recurso. Isso permite que você crie o link do recurso e o recurso compartilhado em qualquer ordem.
- Se você excluir um link de recurso, o recurso compartilhado vinculado não será descartado. Se você descartar um recurso compartilhado, os links de recursos para esse recurso não serão excluídos.
- É possível criar cadeias de links de recursos. No entanto, não há vantagem em fazer isso, pois as APIs seguem apenas o primeiro link de recurso.

Consulte também:

- Conceder permissões nos recursos do Catálogo de Dados

Como criar um link de recurso para uma tabela compartilhada do catálogo de dados

Você pode criar um link de recurso para uma tabela compartilhada em qualquer AWS região usando o AWS Lake Formation console, a API ou AWS Command Line Interface (AWS CLI).

Como criar um link de recurso para a tabela compartilhada (console)

1. Abra o AWS Lake Formation console em https://console.aws.amazon.com/lakeformation/. Faça login como entidade principal que tem a permissão `CREATE_TABLE` do Lake Formation no banco de dados para conter o link do recurso.
2. No painel de navegação, selecione Tabelas em Data Catalog e escolha Criar, Link de recurso.
3. Na página Criar link de recurso, forneça as seguintes informações:

   Nome do link de recurso

   Digite um nome que siga as mesmas regras de um nome de tabela. O nome pode ser o mesmo da tabela compartilhada de destino.

Banco de dados

O banco de dados no catálogo de dados local para conter o link do recurso.

Região do proprietário de tabela compartilhada

Se você estiver criando o link do recurso em uma região diferente, selecione a região da tabela compartilhada de destino.

Tabela compartilhada

Selecione uma tabela compartilhada na lista ou digite um nome de tabela local (de propriedade) ou compartilhada.

A lista contém todas as tabelas compartilhadas com sua conta. Anote o banco de dados e o ID da conta do proprietário listado em cada tabela. Caso você não veja uma tabela que saiba que foi compartilhada com sua conta, verifique o seguinte:

- Se você não for administrador do data lake, verifique se o administrador do data lake concedeu a você as permissões do Lake Formation na tabela.
- Se você for administrador de um data lake e sua conta não estiver na mesma organização da AWS da conta concedente, certifique-se de ter aceitado o convite de compartilhamento de recursos AWS Resource Access Manager (AWS RAM) para o banco de dados. Para obter mais informações, consulte Aceitando um convite de compartilhamento de recursos do AWS RAM.

Banco de dados da tabela compartilhada

Se você selecionou uma tabela compartilhada na lista, esse campo será preenchido com o banco de dados da tabela compartilhada na conta externa. Caso contrário, digite um banco de dados local (para um link de recurso para uma tabela local) ou o banco de dados da tabela compartilhada na conta externa.

Proprietário de tabela compartilhada

Se você selecionou uma tabela compartilhada na lista, esse campo será preenchido com o ID da conta do proprietário da tabela compartilhada. Caso contrário, insira o ID da sua AWS conta (para um link de recurso para uma tabela local) ou o ID da AWS conta que compartilhou a tabela.

4. Selecione Criar para criar o link do recurso.

Em seguida, você pode exibir o nome do link do recurso na coluna Nome na página Tabelas.

5. (Opcional) Conceda a permissão `DESCRIBE` do Lake Formation no link do recurso às entidades principais que devem ser capazes de exibir o link e acessar a tabela de destino.

   No entanto, conceder permissões em um link de recurso não implica que as permissões serão concedidas no banco de dados ou tabela de destino (vinculado). Você deve conceder permissões no banco de dados de destino separadamente para que o table/resource link fique visível no Athena.

Para criar um link de recurso para uma tabela compartilhada na mesma região (AWS CLI)

1. Digite um comando semelhante ao seguinte:

   ```
   aws glue create-table --database-name myissues --table-input
    '{"Name":"my_customers","TargetTable":
   {"CatalogId":"111122223333","DatabaseName":"issues","Name":"customers"}}'
   ```

   Esse comando cria um link de recurso chamado `my_customers` para a tabela compartilhada `customers`, que está no banco de dados `issues` na conta AWS 1111-2222-3333. O link do recurso é armazenado no banco de dados local `myissues`.

2. (Opcional) Conceda a permissão `DESCRIBE` do Lake Formation no link do recurso às entidades principais que devem ser capazes de exibir o link e acessar a tabela de destino.

   No entanto, conceder permissões em um link de recurso não implica que as permissões serão concedidas na tabela de destino (vinculada). Você deve conceder permissões no banco de dados de destino separadamente para que o table/resource link fique visível no Athena.

Para criar um link de recurso para uma tabela compartilhada em uma região diferente (AWS CLI)

1. Digite um comando semelhante ao seguinte:

   ```
   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"
        }
    }
}‘
```

Esse comando cria um link de recurso nomeado `rl_useast1salestb_ireland` na região Europa (Irlanda) para a `useast1_salestb` tabela compartilhada, que está no `useast1_salesdb` banco de dados na conta 444455556666 AWS na região Leste dos EUA (Norte da Virgínia). O link do recurso é armazenado no banco de dados local `ireland_db`.

2. Conceda permissão `DESCRIBE` ao Lake Formation às entidades principais que devem ser capazes de exibir o link e acessar o destino do link por meio do link.

   No entanto, conceder permissões em um link de recurso não implica que as permissões serão concedidas na tabela de destino (vinculada). Você deve conceder permissões na tabela de destino separadamente para que o table/resource link fique visível no Athena.

Consulte também:

- Como os links de recursos funcionam no Lake Formation
- `DESCRIBE`

Como criar um link de recurso para um banco de dados compartilhado do catálogo de dados

Você pode criar um link de recurso para um banco de dados compartilhado usando o AWS Lake Formation console, a API ou AWS Command Line Interface (AWS CLI).

Como criar um link de recurso para um banco de dados compartilhado (console)

1. Abra o AWS Lake Formation console em https://console.aws.amazon.com/lakeformation/. Faça login como administrador de data lake ou como criador de banco de dados.

   Um criador de banco de dados é uma entidade principal que recebeu a permissão `CREATE_DATABASE` do Lake Formation.
2. No painel de navegação, selecione Bancos de dados e escolha Criar, Link de recurso.
3. Na página Criar link de recurso, forneça as seguintes informações:

Nome do link de recurso

Digite um nome que siga as mesmas regras de um nome de banco de dados. O nome pode ser o mesmo do banco de dados compartilhado de destino.

Catálogo de destinos

Selecione o catálogo de destino para o link do recurso do banco de dados.

Região do proprietário do banco de dados compartilhado

Se você estiver criando o link do recurso em uma região diferente, selecione a região do banco de dados compartilhado de destino.

Banco de dados compartilhado

Selecione um banco de dados na lista ou digite um nome de banco de dados local (de propriedade) ou compartilhado.

A lista contém todos os bancos de dados compartilhados com sua conta. Anote o ID da conta do proprietário que está listado em cada banco de dados. Caso você não veja um banco de dados que você sabe que foi compartilhado com sua conta, verifique o seguinte:

- Se você não for administrador do data lake, verifique se o administrador do data lake concedeu a você as permissões do Lake Formation no banco de dados.
- Se você for administrador de um data lake e sua conta não estiver na mesma organização da AWS da conta concedente, certifique-se de ter aceitado o convite de compartilhamento de recursos AWS Resource Access Manager (AWS RAM) para o banco de dados. Para obter mais informações, consulte Aceitando um convite de compartilhamento de recursos do AWS RAM.

Proprietário do banco de dados compartilhado

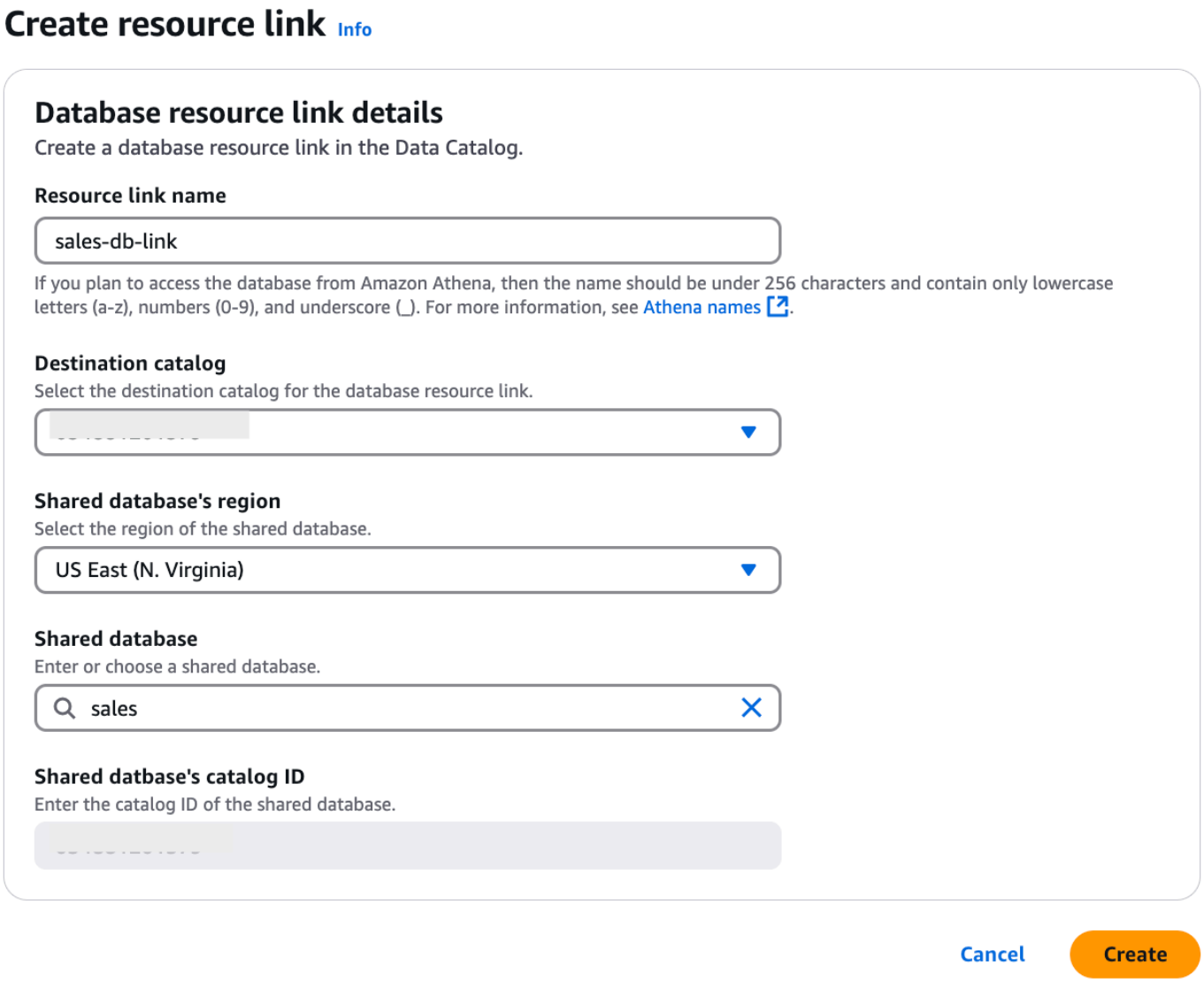
Se você selecionou um banco de dados compartilhado na lista, esse campo será preenchido com o ID da conta do proprietário do banco de dados compartilhado. Caso contrário, insira o ID AWS da sua conta (para um link de recurso para um banco de dados local) ou o ID da AWS conta que compartilhou o banco de dados.

ID do catálogo do banco de dados compartilhado

Insira o ID do catálogo para o banco de dados compartilhado. Ao criar um link de recurso para um banco de dados compartilhado de outra conta da AWS , você precisa especificar

esse ID de catálogo para identificar o Data Catalog da conta que contém o banco de dados de origem.

Quando você seleciona um banco de dados compartilhado no menu suspenso, o sistema preenche automaticamente o ID do catálogo da conta que possui e compartilhou esse banco de dados com você.



4. Selecione Criar para criar o link do recurso.

Em seguida, você pode exibir o nome do link do recurso na coluna Nome na página Bancos de dados.

5. (Opcional) Conceda a permissão `DESCRIBE` do Lake Formation no link do recurso às entidades principais da região da Europa (Irlanda), que devem ser capazes de exibir o link e acessar o banco de dados de destino.

   No entanto, conceder permissões em um link de recurso não implica que as permissões serão concedidas no banco de dados ou tabela de destino (vinculado). Você deve conceder permissões no banco de dados de destino separadamente para que o table/resource link fique visível no Athena.

Para criar um link de recurso para um banco de dados compartilhado na mesma região (AWS CLI)

1. Digite um comando semelhante ao seguinte:

   ```
   aws glue create-database --database-input '{"Name":"myissues","TargetDatabase":
   {"CatalogId":"111122223333","DatabaseName":"issues"}}'
   ```

   Esse comando cria um link de recurso chamado `myissues` para o banco de dados compartilhado`issues`, que está na AWS conta 1111-2222-3333.
2. (Opcional) Conceda a permissão `DESCRIBE` do Lake Formation às entidades principais no link de recurso autorizadas a visualizar o link e acessar o banco de dados ou tabela de destino.

   No entanto, conceder permissões em um link de recurso não implica que as permissões serão concedidas no banco de dados ou tabela de destino (vinculado). Você deve conceder permissões no banco de dados de destino separadamente para que o table/resource link fique visível no Athena.

Para criar um link de recurso para um banco de dados compartilhado em uma região diferente (AWS CLI)

1. Digite um comando semelhante ao seguinte:

   ```
   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"
        }
    }
}'
```

Esse comando cria um link de recurso chamado `rl_useast1shared_irelanddb` na AWS conta 111122223333 na região Europa (Irlanda) para o banco de dados compartilhado`useast1shared_db`, que está na AWS conta 444455556666 na região Leste dos EUA (Norte da Virgínia).

2. Conceda a permissão `DESCRIBE` do Lake Formation às entidades principais da região da Europa (Irlanda), que devem ser capazes de exibir o link e acessar o destino do link por meio do link.

> Consulte também:
>
> - Como os links de recursos funcionam no Lake Formation
> - `DESCRIBE`

Manipulação de links de recursos em AWS Glue APIs

As tabelas a seguir explicam como as APIs do catálogo de dados do AWS Glue lidam com links de recursos de banco de dados e tabelas. Para todas as operações de API do `Get*`, somente bancos de dados e tabelas nos quais o chamador tem permissões são retornados. Além disso, ao acessar um banco de dados ou tabela de destino por meio de um link de recurso, você deve ter as permissões AWS Identity and Access Management (IAM) e Lake Formation no link de destino e no link do recurso. A permissão do Lake Formation exigida nos links de recursos é `DESCRIBE`. Para obter mais informações, consulte `DESCRIBE`.

Operações de API de banco de dados

Operação de API	Gestão de links de recursos
`CreateDatabase`	Se o banco de dados for um link de recurso, ele cria o link de recurso para o banco de dados de destino designado.
`UpdateDatabase`	Se o banco de dados designado for um link de recurso, ele segue o link e atualiza o banco de dados de destino. Se o link de recurso

Operação de API	Gestão de links de recursos
	precisar ser modificado para vincular a um banco de dados diferente, você deverá excluí-lo e criar um novo.
`DeleteDatabase`	Exclua o link do recurso. Isto não exclui o banco de dados vinculado (de destino).
`GetDatabase`	Se o chamador tiver permissões no alvo, ele segue o link para retornar as propriedades do alvo. Caso contrário, ele retornará as propriedades do link.
`GetDatabases`	Retorna uma lista de bancos de dados, incluindo links de recursos. Para cada link de recurso no conjunto de resultados, a operação segue o link para obter as propriedades do link de destino. Você deve especificar `ResourceShareType = ALL` para ver os bancos de dados compartilhados com sua conta.

Operações de API de tabela

Operação de API	Gestão de links de recursos
`CreateTable`	Se o banco de dados for um link de recurso, ele segue o link do banco de dados e cria uma tabela no banco de dados de destino. Se a tabela for um link de recurso, a operação criará o link de recurso no banco de dados designado. Não há suporte para criar um link de recurso de tabela por meio de um link de recurso de banco de dados.
`UpdateTable`	Se a tabela ou o banco de dados designado for um link de recurso, isso atualiza a tabela de destino. Se a tabela e o banco de dados forem links de recursos, a operação falhará.
`DeleteTable`	Se o banco de dados designado for um link de recurso, ele segue o link e exclui a tabela ou o link do recurso da tabela no banco de dados de destino. Se a tabela for um link de recurso, a operação excluirá o link de recurso da tabela no banco de dados designado

Operação de API	Gestão de links de recursos
	. A exclusão de um link de recurso de tabela não exclui a tabela de destino.
`BatchDeleteTable`	Igual a `DeleteTable` .
`GetTable`	Se o banco de dados designado for um link de recurso, ele segue o link do banco de dados e retorna a tabela ou o link do recurso da tabela do banco de dados de destino. Caso contrário, se a tabela for um link de recurso, a operação seguirá o link e retornará as propriedades da tabela de destino.
`GetTables`	Se o banco de dados designado for um link de recurso, ele segue o link do banco de dados e retorna as tabelas e os links de recursos da tabela do banco de dados de destino. Se o banco de dados de destino for um banco de dados compartilhado de outra AWS conta, a operação retornará somente as tabelas compartilhadas nesse banco de dados. Ele não segue os links de recursos da tabela no banco de dados de destino. Caso contrário, se o banco de dados designado for um banco de dados local (de propriedade), a operação retornará todas as tabelas no banco de dados local e seguirá cada link de recurso da tabela para retornar as proprieda des da tabela de destino.
`SearchTables`	Retorna tabelas e links de recursos de tabelas. Ele não segue links para retornar as propriedades da tabela de destino. Você deve especificar `ResourceShareType = ALL` para ver as tabelas compartilhadas com sua conta.
`GetTableVersion`	Igual a `GetTable`.
`GetTableVersions`	Igual a `GetTable`.
`DeleteTableVersion`	Igual a `DeleteTable` .
`BatchDeleteTableVersion`	Igual a `DeleteTable` .

Operações de API de partição

Operação de API	Gestão de links de recursos
`CreatePartition`	Se o banco de dados designado for um link de recurso, ele segue o link do banco de dados e cria uma partição na tabela designada no banco de dados de destino. Se a tabela for um link de recurso, a operação segue o link do recurso e cria a partição na tabela de destino. A criação de uma partição por meio de um link de recurso de tabela e de um link de recurso de banco de dados não é suportada.
`BatchCreatePartition`	Igual a `CreatePartition` .
`UpdatePartition`	Se o banco de dados designado for um link de recurso, ele segue o link do banco de dados e atualiza a partição na tabela designada no banco de dados de destino. Se a tabela for um link de recurso, a operação segue o link do recurso e atualiza a partição na tabela de destino. A atualização de uma partição por meio de um link de recurso de tabela e de um link de recurso de banco de dados não é suportada.
`DeletePartition`	Se o banco de dados designado for um link de recurso, ele segue o link do banco de dados e exclui a partição na tabela designada no banco de dados de destino. Se a tabela for um link de recurso, a operação segue o link do recurso e exclui a partição na tabela de destino. A exclusão de uma partição por meio de um link de recurso de tabela e de um link de recurso de banco de dados não é suportada.
`BatchDeletePartition`	Igual a `DeletePartition` .
`GetPartition`	Se o banco de dados designado for um link de recurso, ele segue o link do banco de dados e retorna as informações da partição da tabela designada. Caso contrário, se a tabela for um link de recurso, a operação seguirá o link e retornará as informações da

Operação de API	Gestão de links de recursos
	partição. Se a tabela e o banco de dados forem links de recursos, ele retornará um conjunto de resultados vazio.
`GetPartitions`	Se o banco de dados designado for um link de recurso, ele segue o link do banco de dados e retorna as informações de partição de todas as partições na tabela designada. Caso contrário, se a tabela for um link de recurso, a operação seguirá o link e retornará as informações da partição. Se a tabela e o banco de dados forem links de recursos, ele retornará um conjunto de resultados vazio.
`BatchGetPartition`	Igual a `GetPartition` .

User-defined funções: operações de API

Operação de API	Gestão de links de recursos
(Todas as operações de API)	Se o banco de dados for um link de recurso, ele segue o link do recurso e executa a operação no banco de dados de destino.

Consulte também:

- Como os links de recursos funcionam no Lake Formation

Acessar tabelas entre regiões

O Lake Formation suporta a consulta de tabelas do Catálogo de Dados em todas as AWS regiões. Você pode acessar dados em uma região de outras regiões usando o Amazon Athena, o Amazon EMR e o AWS Glue ETL criando links de recursos em outras regiões apontando para os bancos de dados e tabelas de origem. Com o acesso à tabela entre regiões, você pode acessar dados entre regiões sem copiar os dados subjacentes ou os metadados no catálogo de dados.

Por exemplo, você pode compartilhar um banco de dados ou uma tabela em uma conta de produtor com uma conta de consumidor na Região A. Após aceitar o convite de compartilhamento de recursos na Região A, o administrador do data lake da conta do consumidor pode criar links de recursos

para o recurso compartilhado na Região B. O administrador da conta do consumidor pode conceder permissões sobre o recurso compartilhado com as entidades principais do IAM nessa conta na Região A, e conceder permissões de link de recurso na Região B. Ao usar o link do recurso, as entidades principais da conta do consumidor podem consultar os dados compartilhados da Região B.

Você também ser host da fonte de dados do Amazon S3 na Região A em uma conta de produtor, e registrar o local dos dados em uma conta central na Região B. Você pode criar recursos do catálogo de dados na conta central, configurar permissões do Lake Formation, e compartilhar dados com consumidores em sua conta ou com contas externas na Região B. O atributo entre regiões permite que os usuários acessem essas tabelas do catálogo de dados da Região C usando links de recursos.

Ao usar esse atributo, você pode consultar bancos de dados federados em Apache Hive repositórios entre regiões e também unir tabelas na região local com tabelas em outra região ao executar consultas.

O Lake Formation oferece suporte aos seguintes recursos com acesso a tabelas entre regiões:

- LF-Tag controle de acesso baseado
- Fine-grained permissões de controle de acesso
- Operações de gravação no banco de dados compartilhado ou na tabela com as permissões apropriadas
- Cross-account compartilhamento de dados no nível da conta e direto com os diretores do IAM

Non-administrative usuários com `Create_Table` permissões `Create_Database` e permissões podem criar links de recursos entre regiões.

> **Note**
>
> Você pode criar links de recursos entre regiões em qualquer região e acessar dados sem aplicar as permissões do Lake Formation. Para dados de origem no Amazon S3 que não estão registrados no Lake Formation, o acesso é determinado pelas políticas de permissões do IAM para o Amazon AWS Glue S3 e pelas ações.

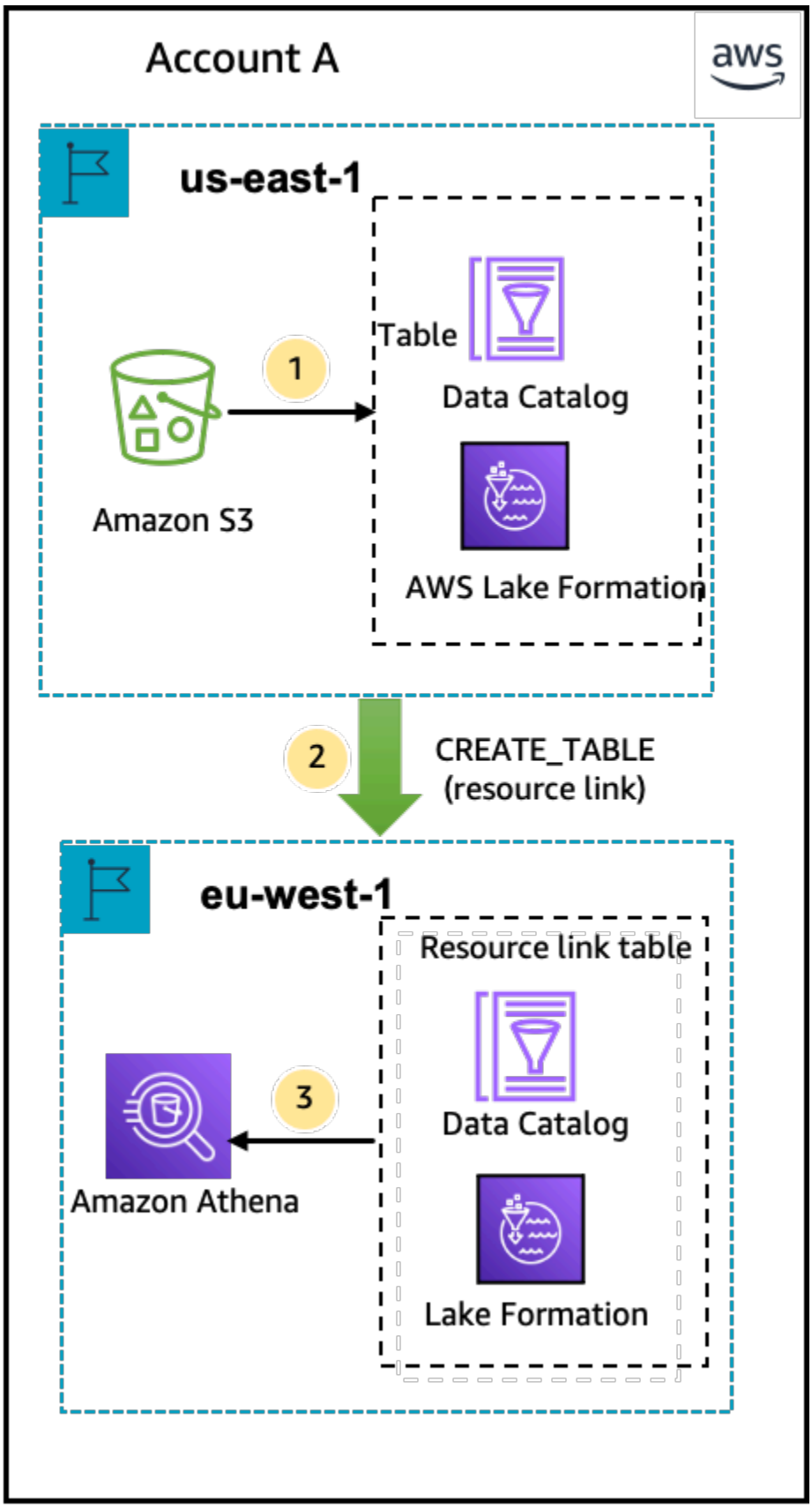
Para conhecer as limitações, consulte Limitações de acesso aos dados entre regiões.

Fluxos de trabalho

Os diagramas a seguir mostram os fluxos de trabalho para acessar dados entre AWS regiões a partir da mesma AWS conta e de uma conta externa.

Fluxo de trabalho para acessar tabelas compartilhadas dentro da mesma AWS account

No diagrama abaixo, os dados são compartilhados com um usuário na mesma AWS conta na região Leste dos EUA (Norte da Virgínia), e o usuário consulta os dados compartilhados da região Europa (Irlanda).


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

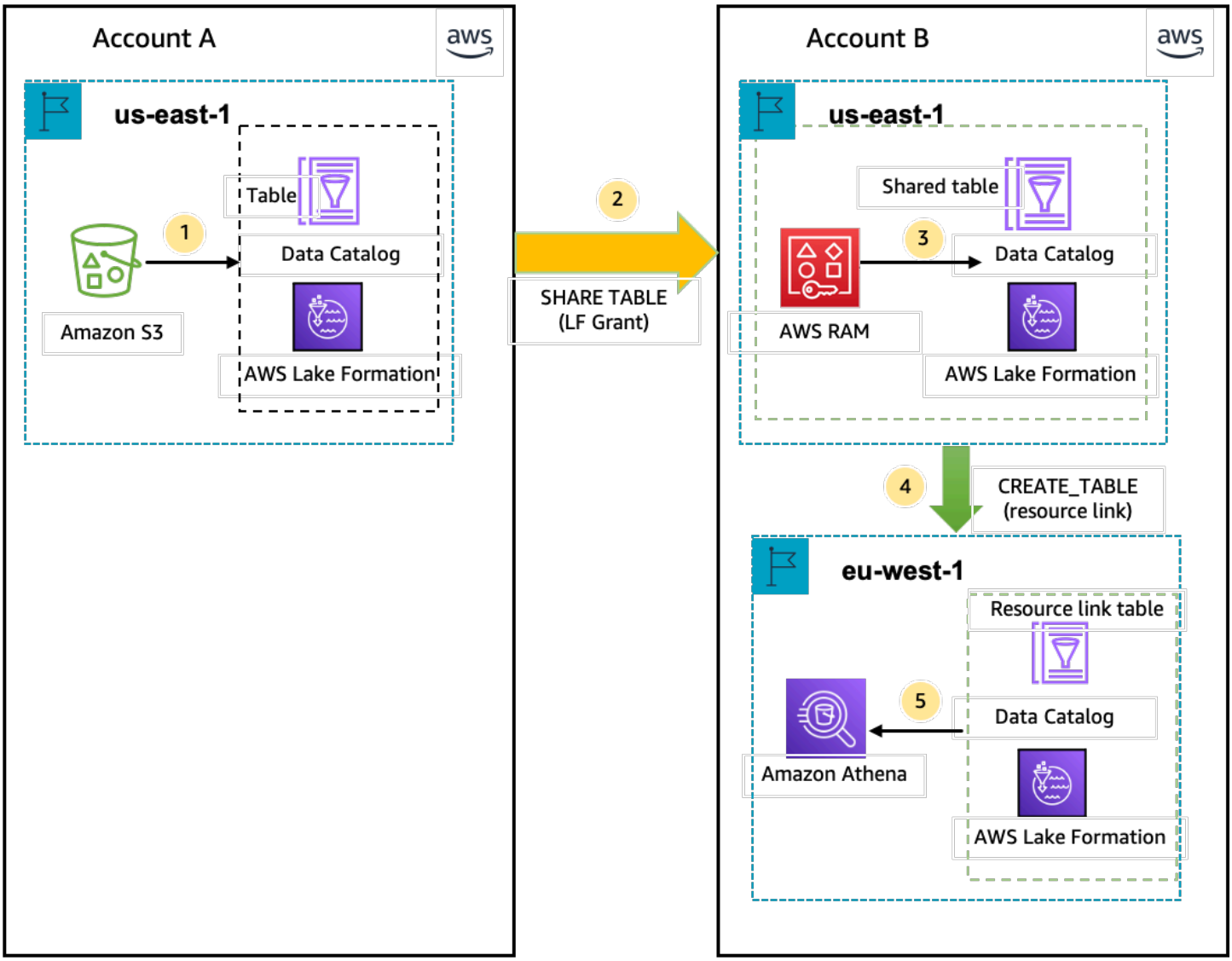
O administrador do data lake executa as seguintes atividades (etapas 1 e 2):

1. Um administrador de data lake configura uma AWS conta com os bancos de dados e tabelas do Catálogo de Dados e registra uma localização de dados do Amazon S3 no Lake Formation na região Leste dos EUA (Norte da Virgínia).

   Conceder a permissão `Select` em um recurso do catálogo de dados (tabela de produtos no diagrama) a uma entidade principal (usuário) na mesma conta.
2. Cria um link de recurso na região da Europa (Irlanda) apontando para a tabela de origem na região Leste dos EUA (Norte da Virgínia). Conceder a permissão `DESCRIBE` no link do recurso da região da Europa (Irlanda) à entidade principal.
3. O usuário consulta a tabela da região da Europa (Irlanda) usando Athena.

Fluxo de trabalho para acessar tabelas compartilhadas com um externo AWS account

No diagrama abaixo, a conta do produtor (Conta A) hospeda o bucket Amazon S3, registra o local dos dados e compartilha uma tabela do catálogo de dados com uma conta de consumidor (Conta B) na região Leste dos EUA (Norte da Virgínia) e um usuário da conta do consumidor (Conta B) consulta a tabela da região Europa (Irlanda).



1. Um administrador do data lake configura uma AWS conta (conta do produtor) com os recursos do catálogo de dados e um local de dados do Amazon S3 registrado no Lake Formation na região Leste dos EUA (Norte da Virgínia).
2. O administrador do data lake da conta do produtor compartilha uma tabela do catálogo de dados com uma conta de consumidor.
3. O administrador do data lake da conta do consumidor aceita o convite de compartilhamento de dados na região Leste dos EUA (Norte da Virgínia) e concede a permissão `Select` na tabela compartilhada a uma entidade principal da mesma região.
4. O administrador do data lake da conta do consumidor cria um link de recurso na região da Europa (Irlanda) apontando para a tabela compartilhada de destino na região Leste dos EUA (Norte

da Virgínia) e concede ao usuário a permissão `DESCRIBE` no link do recurso da região Europa (Irlanda).

5. O usuário consulta os dados da região da Europa (Irlanda) usando o Athena.

Como configurar o acesso à tabela entre regiões

Para acessar dados de uma região diferente, você precisa primeiro configurar os bancos de dados e as tabelas do catálogo de dados na região em que você registra suo local de dados do Amazon S3. Você pode compartilhar os bancos de dados e tabelas do catálogo de dados com as entidades principais na sua conta ou em outra. Em seguida, você precisa criar administradores de data lake que possam criar links de recursos apontando para o local de destino dos dados compartilhados nas regiões onde os usuários consultam os dados.

Como consultar dados compartilhados na mesma conta de uma região diferente

Nesta seção, a região da tabela compartilhada de destino é chamada de Região A e os usuários executam consultas na Região B.

1. Configuração da conta na Região A (onde você cria e compartilha os dados)

   O administrador do data lake precisa concluir as ações a seguir:

   a. Registre um local de dados do Amazon S3.

      Para obter mais informações, consulte Adicionar uma localização do Amazon S3 ao seu data lake.

   b. Crie bancos de dados e tabelas na conta. Isso também pode ser feito por um usuário não administrativo que tenha permissões para criar bancos de dados e tabelas.

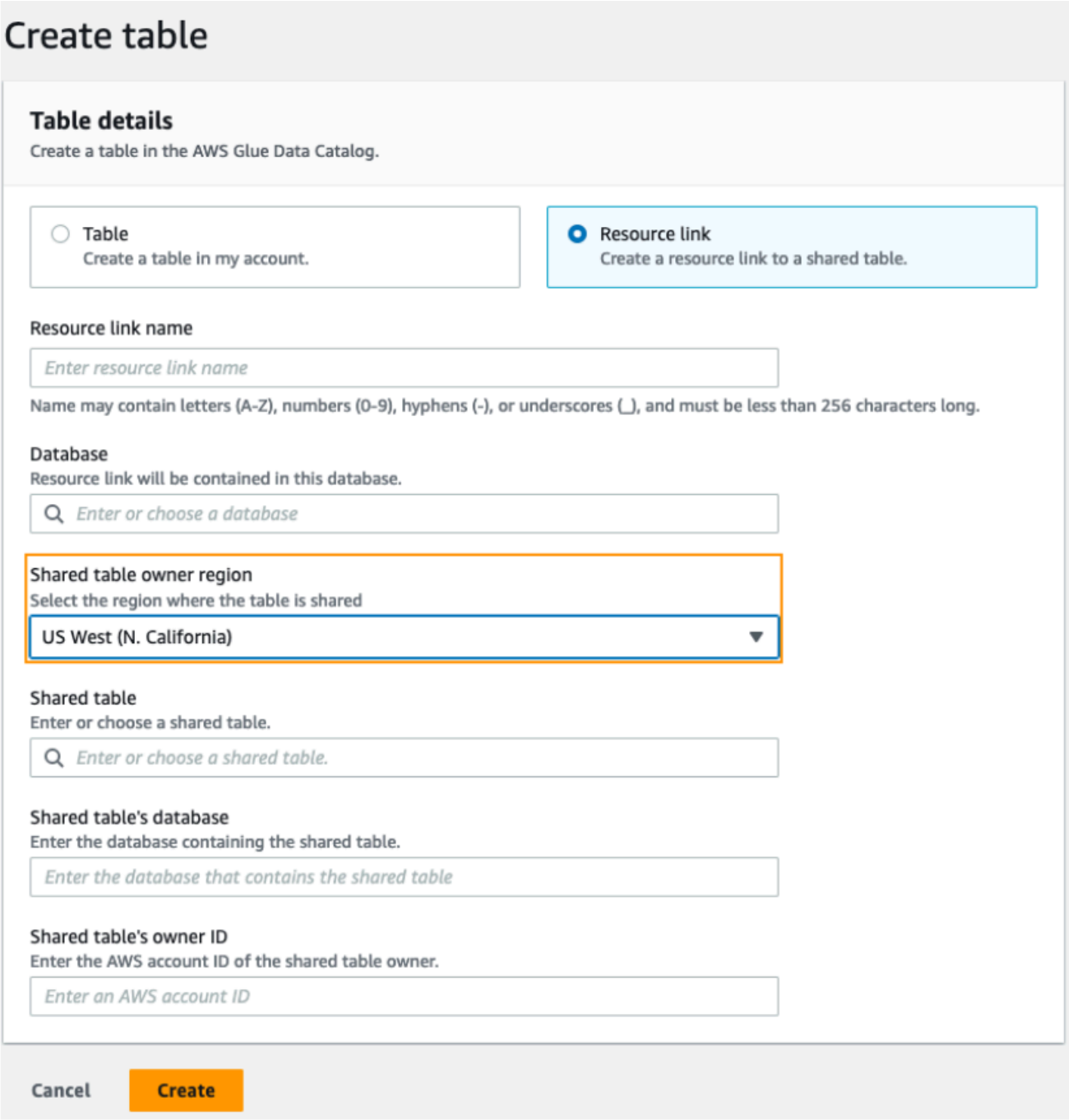
   c. Conceda permissões de dados em uma tabela às entidades principais com `Grantable permissions`.

      Para ter mais informações, consulte Conceder permissões nos recursos do Catálogo de Dados.

2. Configuração da conta na Região B (onde você acessa os dados)

   O administrador do data lake precisa concluir as ações a seguir:

   a. Crie um link de recurso na Região B apontando para a tabela compartilhada de destino na Região A. Especifique a Região proprietária da tabela compartilhada na tela Criar tabela.



Para obter instruções sobre como criar links de recursos para bancos de dados e tabelas, consulte Criação de links de recursos.

b. Conceda a permissão `Describe` às entidades principais do IAM no link do recurso na Região B.

Para obter mais informações sobre a concessão de permissões em links de recursos, consulte Como conceder permissões de links de recursos.

As entidades principais do IAM na Região B podem consultar a tabela de destino por meio do link usando o Athena.

Como acessar dados de várias contas de uma região diferente

1. Producer/grantor configuração da conta

   O administrador do data lake precisa concluir as ações a seguir:

   a. Configure a producer/grantor conta na Região A.
   b. Registre um local de dados do Amazon S3 na Região A.
   c. Criar bancos de dados e tabelas. Isso pode ser feito por um usuário não administrativo que tenha permissões para criar tabelas.
   d. Conceda permissões de dados para a consumer/grantee conta em uma tabela na Região A com`Grantable permissions`.

      Para obter mais informações, consulte Compartilhando tabelas e bancos de dados do Catálogo de Dados entre Contas da AWS ou diretores do IAM de contas externas.

2. Consumer/grantee configuração da conta

   O administrador do data lake precisa concluir as ações a seguir:

   a. Aceite o convite de compartilhamento de recursos AWS RAM da Região A.
   b. Crie um link de recurso na Região B apontando para a tabela compartilhada. A região B é onde os usuários desejarão consultar a tabela.
   c. Conceda permissões de dados na tabela compartilhada às entidades principais do IAM na Região A.

      > **Note**
      >
      > Você deve conceder permissões para a tabela compartilhada na mesma região em que a tabela foi compartilhada.

   d. Conceda permissões às entidades principais no link do recurso na Região B.

      As entidades principais da conta do consumidor na Região B então consultam a tabela compartilhada da Região B usando o Athena.

Segurança em AWS Lake Formation

A segurança na nuvem AWS é a maior prioridade. Como AWS cliente, você se beneficia de uma arquitetura de data center e rede criada para atender aos requisitos das organizações mais sensíveis à segurança.

A segurança é uma responsabilidade compartilhada entre você AWS e você. O modelo de responsabilidade compartilhada descreve isto como segurança da nuvem e segurança na nuvem.

- Segurança da nuvem — AWS é responsável por proteger a infraestrutura que executa AWS os serviços na AWS nuvem. AWS também fornece serviços que você pode usar com segurança. Auditores de terceiros testam e verificam regularmente a eficácia da nossa segurança como parte dos programas de conformidade da AWS. Para saber mais sobre os programas de conformidade que se aplicam ao AWS Lake Formation, consulte AWS Services in Scope by Compliance Program.
- Segurança na nuvem — Sua responsabilidade é determinada pelo AWS serviço que você usa. Você também é responsável por outros fatores, incluindo a confidencialidade de seus dados, os requisitos da empresa e as leis e regulamentos aplicáveis.

Esta documentação ajuda você a entender como aplicar o modelo de responsabilidade compartilhada ao usar o Lake Formation. Os tópicos a seguir mostram como configurar o Lake Formation para atender aos seus objetivos de segurança e conformidade. Você também aprenderá a usar outros AWS serviços que ajudam a monitorar e proteger seus recursos do Lake Formation.

Tópicos

- Proteção de dados no Lake Formation
- Segurança de infraestrutura em AWS Lake Formation
- Prevenção contra o ataque do “substituto confuso” em todos os serviços
- Registro de eventos de segurança em AWS Lake Formation

Proteção de dados no Lake Formation

O modelo de responsabilidade compartilhada da AWS se aplica à proteção de dados no . Conforme descrito neste modelo, AWS é responsável por proteger a infraestrutura global que executa todos os Nuvem AWS. Você é responsável por manter o controle sobre o conteúdo hospedado nessa

infraestrutura. Você também é responsável pelas tarefas de configuração e gerenciamento de segurança dos Serviços da AWS que usa. Para obter mais informações sobre privacidade de dados, consulte Perguntas frequentes sobre privacidade de dados . Para obter informações sobre proteção de dados na Europa, consulte o Centro de Regulamento Geral sobre a Proteção de Dados (RGPD).

Para fins de proteção de dados, recomendamos que você proteja Conta da AWS as credenciais e configure usuários individuais com Centro de Identidade do AWS IAM ou AWS Identity and Access Management (IAM). Dessa maneira, cada usuário receberá apenas as permissões necessárias para cumprir suas obrigações de trabalho. Recomendamos também que você proteja seus dados das seguintes formas:

- Use uma autenticação multifator (MFA) com cada conta.
- Use SSL/TLS para se comunicar com AWS os recursos. Exigimos TLS 1.2 e recomendamos TLS 1.3.
- Configure a API e o registro de atividades do usuário com AWS CloudTrail. Para obter informações sobre o uso de CloudTrail trilhas para capturar AWS atividades, consulte Como trabalhar com CloudTrail trilhas no Guia AWS CloudTrail do usuário.
- Use soluções de AWS criptografia, juntamente com todos os controles de segurança padrão Serviços da AWS.
- Use serviços gerenciados de segurança avançada, como o Amazon Macie, que ajuda a localizar e proteger dados sensíveis armazenados no Amazon S3.
- Se você precisar de módulos criptográficos validados pelo FIPS 140-3 ao acessar AWS por meio de uma interface de linha de comando ou de uma API, use um endpoint FIPS. Para saber mais sobre os endpoints FIPS disponíveis, consulte Federal Information Processing Standard (FIPS) 140-3.

É altamente recomendável que nunca sejam colocadas informações confidenciais ou sensíveis, como endereços de e-mail de clientes, em tags ou campos de formato livre, como um campo Nome. Isso inclui quando você trabalha com ou Serviços da AWS usa o console, a API ou AWS os SDKs. AWS CLI Quaisquer dados inseridos em tags ou em campos de texto de formato livre usados para nomes podem ser usados para logs de faturamento ou de diagnóstico. Se você fornecer um URL para um servidor externo, é fortemente recomendável que não sejam incluídas informações de credenciais no URL para validar a solicitação nesse servidor.

Criptografia em repouso

AWS O Lake Formation oferece suporte à criptografia de dados nas seguintes áreas:

- Dados no data lake do Amazon Simple Storage Service (Amazon S3).

  O Lake Formation oferece suporte à criptografia de dados com AWS Key Management Service (AWS KMS). Os dados geralmente são gravados no data lake por meio de trabalhos de extração, transformação e carregamento (ETL) do AWS Glue. Para obter informações sobre como criptografar dados gravados por trabalhos do AWS Glue, consulte Criptografar dados gravados por crawlers, trabalhos e endpoints de desenvolvimento no Guia do desenvolvedor do AWS Glue .
- O AWS Glue Data Catalog, que é onde o Lake Formation armazena tabelas de metadados que descrevem os dados no data lake.

  Para obter mais informações, consulte Criptografando seu catálogo de dados no Guia do desenvolvedor do AWS Glue .

Para adicionar um local do Amazon S3 como armazenamento em seu data lake, você registra o local no AWS Lake Formation. Em seguida, você pode usar as permissões do Lake Formation para um controle de acesso refinado aos objetos AWS Glue Data Catalog que apontam para esse local e aos dados subjacentes no local.

O Lake Formation suporta o registro de uma localização do Amazon S3 que contém dados criptografados. Para obter mais informações, consulte Registrando uma localização criptografada do Amazon S3.

Segurança de infraestrutura em AWS Lake Formation

Como um serviço gerenciado, o AWS Lake Formation é protegido pelos procedimentos AWS globais de segurança de rede descritos no whitepaper Amazon Web Services: Overview of Security Processes.

Você usa chamadas de API AWS publicadas para acessar o Lake Formation pela rede. Os clientes devem oferecer suporte a Transport Layer Security (TLS) 1.0 ou posterior. Recomendamos TLS 1.2 ou posterior. Os clientes também devem ter compatibilidade com conjuntos de criptografia com perfect forward secrecy (PFS) como Ephemeral Diffie-Hellman (DHE) ou Ephemeral Elliptic Curve Diffie-Hellman (ECDHE). A maioria dos sistemas modernos como Java 7 e versões posteriores oferece compatibilidade com esses modos.

Além disso, as solicitações devem ser assinadas usando um ID da chave de acesso e uma chave de acesso secreta associada a uma entidade principal do IAM. Ou você pode usar o AWS

Security Token Service (AWS STS) para gerar credenciais de segurança temporárias para assinar solicitações.

Prevenção contra o ataque do “substituto confuso” em todos os serviços

“Confused deputy” é um problema de segurança no qual uma entidade sem permissão para executar uma ação pode coagir uma entidade mais privilegiada a executá-la. Em AWS, a falsificação de identidade entre serviços pode resultar no problema confuso do deputado. A personificação entre serviços pode ocorrer quando um serviço (o serviço de chamada) chama outro serviço (o serviço chamado). O serviço de chamada pode ser manipulado de modo a usar suas permissões para atuar nos recursos de outro cliente de uma forma na qual ele não deveria ter permissão para acessar. Para evitar isso, AWS fornece ferramentas que ajudam você a proteger seus dados para todos os serviços com diretores de serviços que receberam acesso aos recursos em sua conta.

Recomendamos o uso das chaves de contexto de condição global `aws:SourceArn` e `aws:SourceAccount` em políticas de recursos para limitar as permissões de acesso ao recurso que o AWS Lake Formation concede a outro serviço. Se você usar as duas chaves de contexto de condição global, o valor `aws:SourceAccount` e a conta no valor `aws:SourceArn` deverão usar o mesmo ID de conta quando usados na mesma declaração de política.

Atualmente, o Lake Formation só é compatível `aws:SourceArn` com o seguinte formato:

```
arn:aws:lakeformation:aws-region:account-id:*
```

O exemplo a seguir mostra como você pode usar as teclas de contexto de condição global `aws:SourceArn` e `aws:SourceAccount` no Lake Formation para evitar o problema do "substituto confuso".

JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Sid": "ConfusedDeputyPreventionExamplePolicy",
      "Effect": "Allow",
      "Principal": {
```

```
        "Service": "lakeformation.amazonaws.com"
      },
      "Action": [
        "sts:AssumeRole"
      ],
      "Condition": {
        "StringEquals": {
          "aws:SourceAccount": "123456789012"
        },
        "ArnEquals": {
          "aws:SourceArn": "arn:aws:lakeformation:aws-region:123456789012:*"
        }
      }
    }
  ]
}
```

Registro de eventos de segurança em AWS Lake Formation

AWS O Lake Formation é integrado com AWS CloudTrail, um serviço que fornece um registro das ações realizadas por um usuário, função ou AWS serviço no Lake Formation. CloudTrail captura todas as chamadas de API para Lake Formation como eventos. As chamadas capturadas incluem chamadas do console do Lake Formation, do AWS Command Line Interface, e chamadas de código para as operações da API do Lake Formation.

Para obter mais informações sobre o log de eventos no Lake Formation, consulte Registrando chamadas da API AWS Lake Formation usando AWS CloudTrail.

> **Note**
>
> `GetTableObjects`, `UpdateTableObjects` e `GetWorkUnitResults` são operações de plano de dados de alto volume. As chamadas para eles não APIs estão registradas no momento. CloudTrail Para obter mais informações sobre as operações do plano de dados em CloudTrail, consulte Registro de eventos de dados para trilhas no Guia AWS CloudTrail do usuário.
>
> Mudanças no Lake Formation para apoiar CloudTrail eventos adicionais serão documentadas emHistórico do documento para AWS Lake Formation.

Integrando serviços de terceiros com Lake Formation

A integração com o AWS Lake Formation permite que serviços de terceiros acessem com segurança dados em seus data lakes baseados no Amazon S3. Você pode usar o Lake Formation como seu mecanismo de autorização para gerenciar ou aplicar permissões ao seu data lake com AWS serviços integrados, como AWS Glue ETL, Amazon Athena, Amazon EMR e Redshift Spectrum. O Lake Formation oferece duas opções para integrar serviços:

1. As configurações de integração do aplicativo Lake Formation: O Lake Formation pode vender credenciais temporárias com escopo reduzido na forma de tokens AWS STS para locais registrados do Amazon S3 com base nas permissões efetivas, para que aplicativos autorizados possam acessar dados em nome dos usuários.
2. Aplicação central: as operações de API de consulta do Lake Formation recuperam dados do Amazon S3 e filtram os resultados com base nas permissões efetivas. O mecanismo ou aplicativo que se integra à operação da API de consulta pode depender do Lake Formation para avaliar as permissões da identidade de chamada e filtrar com segurança os dados com base nessas permissões. Third-party os mecanismos de consulta só veem e operam com dados filtrados.

A venda de credenciais do Lake Formation não se integra às consultas spark sql. A venda de credenciais só funciona com consultas executadas na biblioteca ETL. AWS Glue

Tópicos

- Como usar a integração de aplicativos Lake Formation

Como usar a integração de aplicativos Lake Formation

O Lake Formation permite que serviços terceirizados se integrem ao Lake Formation e obtenham acesso temporário aos dados do Amazon S3 em nome de seus usuários por meio do uso GetTemporaryGlueTableCredentialse GetTemporaryGluePartitionCredentialsdas operações. Isso permite que serviços de terceiros usem o mesmo recurso de autorização e venda de credenciais usado pelos demais serviços de AWS análise. Esta seção descreve como usar essas operações de API para integrar um mecanismo de consulta de terceiros com o Lake Formation.

Por padrão, essas operações de API estão desativadas. Há duas opções para autorizar a integração de aplicativos com o Lake Formation:

- Configure tags de sessão do IAM que são validadas sempre que as operações da API de integração de aplicativos são chamadas

  Para obter mais informações, consulte Como habilitar permissões para que um mecanismo de consulta de terceiros chame operações de API de integração de aplicativos.

- Ative a opção que permite que mecanismos externos acessem dados em locais do Amazon S3 com acesso total à tabela.

  Essa opção permite que mecanismos de consulta e aplicativos obtenham credenciais sem tags de sessão do IAM se o usuário tiver acesso total à tabela. Ele fornece benefícios de desempenho para mecanismos de consulta e aplicativos, além de simplificar o acesso aos dados. O Amazon EMR no Amazon EC2 é capaz de utilizar essa configuração.

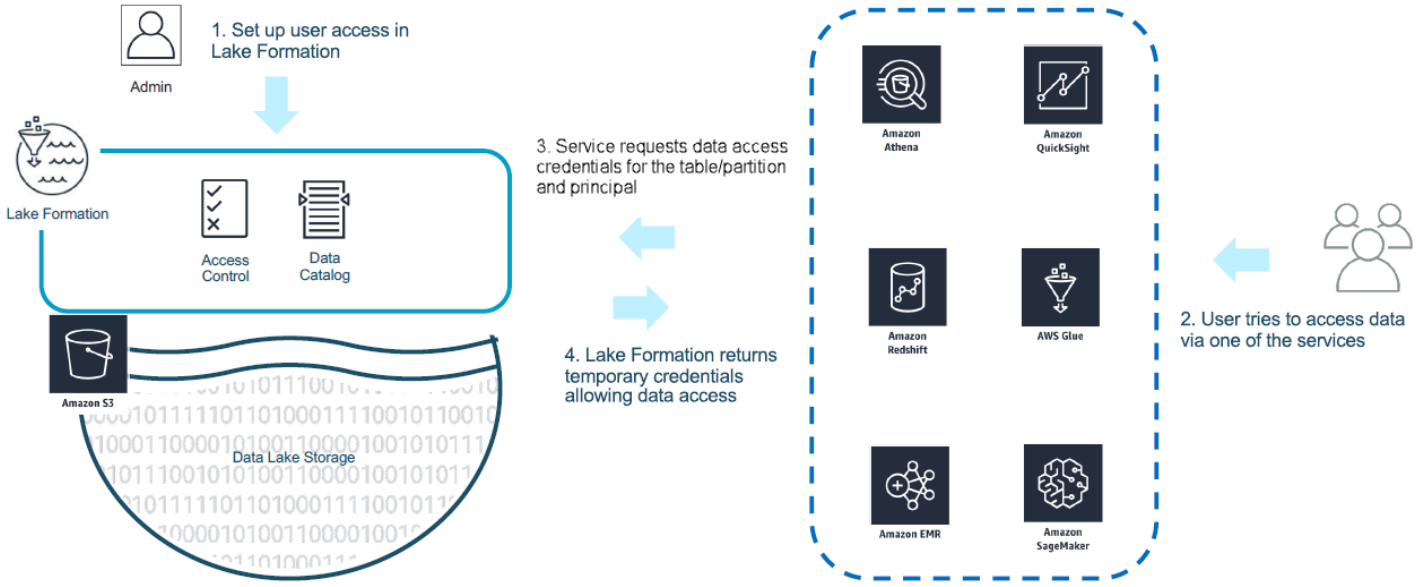
  Para obter mais informações, consulte Integração de aplicativos para acesso total à tabela.

Tópicos

- Como funciona a integração de aplicações do Lake Formation
- Perfis e responsabilidades na integração de aplicativos do Lake Formation
- Lake Formation fluxo de trabalho para operações de API de integração de aplicativos
- Como registrar um mecanismo de consulta de terceiros
- Como habilitar permissões para que um mecanismo de consulta de terceiros chame operações de API de integração de aplicativos
- Integração de aplicativos para acesso total à tabela

Como funciona a integração de aplicações do Lake Formation

Esta seção descreve como usar as operações da API de integração de aplicativos para integrar um aplicativo de terceiros (mecanismo de consulta) com o Lake Formation.



1. O administrador do Lake Formation executa as seguintes atividades:
   - Registra um local do Amazon S3 no Lake Formation fornecendo um perfil do IAM (usado para fornecimento de credenciais) que tem permissões apropriadas para acessar dados dentro do local do Amazon S3
   - Registra um aplicativo de terceiros para poder chamar as operações de API de fornecimento credenciais do Lake Formation. Consulte the section called “Como registrar um mecanismo de consulta de terceiros”
   - Concede permissões para que os usuários acessem bancos de dados e tabelas

     Por exemplo, se você quiser publicar um conjunto de dados de sessões de usuário que inclua algumas colunas contendo informações de identificação pessoal (PII), para restringir o acesso, atribua a essas colunas uma LF-TBACtag chamada “classificação” com o valor “sensível”. Em seguida, defina uma permissão para um analista de negócios acessar os dados das sessões do usuário, mas exclui as colunas marcadas com classificação = sensível.
2. Uma entidade principal (usuário) envia uma consulta para um serviço integrado.
3. O aplicativo integrado solicita ao Lake Formation informações e credenciais da tabela para acessar a tabela.
4. Se a entidade principal que faz a consulta estiver autorizada a acessar a tabela, o Lake Formation retornará as credenciais para o aplicativo integrado, que permite o acesso aos dados.

> **Note**
>
> O Lake Formation não acessa os dados subjacentes ao fornecer credenciais.

5. O serviço integrado lê dados do Amazon S3, filtra as colunas com base nas políticas recebidas e retorna os resultados à entidade principal.

> **Important**
>
> As operações de API do Lake Formation de fornecimento de credenciais permitem um modelo de fiscalização distribuída com negação explícita em caso de falha (fail-close). Isso introduz um modelo de segurança tripartido entre clientes, serviços terceirizados e Lake Formation. Os serviços integrados são confiáveis para aplicar adequadamente as permissões do Lake Formation (fiscalização distribuída).

O serviço integrado é responsável por filtrar os dados lidos do Amazon S3 com base nas políticas retornadas do Lake Formation antes que os dados filtrados sejam devolvidos ao usuário. Os serviços integrados seguem um modelo de "fail-close", o que significa que eles devem falhar na consulta se não conseguirem aplicar as permissões do Lake Formation necessárias.

Perfis e responsabilidades na integração de aplicativos do Lake Formation

A seguir estão as funções e suas responsabilidades associadas para permitir a integração de aplicativos de terceiros com AWS Lake Formation.

Perfil	Responsabilidade
O consumidor	• Ative a configuração de integração do aplicativo Lake Formation (consulte the section called “Como registrar um mecanismo de consulta de terceiros”). • Registra explicitamente terceiros aprovados no Lake Formation (consulte the section called “Como registrar um mecanismo de consulta de terceiros”). • Testa e valida soluções de terceiros com as permissões do Lake Formation.

Perfil	Responsabilidade
	• Monitora e audita o uso de terceiros das operações da API de fornecimento automática de credenciais do Lake Formation.
Terceiros	• Documenta publicamente o recurso suportado para cada revisão de software e fornece instruções para ativá-lo corretamente. • Anuncia com precisão os recursos suportados ao chamar as operações da API de fornecimento de credenciais do Lake Formation (de acordo com a documentação). • Armazena e gerencia com segurança as credenciais fornecidas para evitar vazamentos de credenciais e aumento de privilégios. • Impõe permissões com base nos recursos suportados e retorna somente dados filtrados aos usuários • Falha na consulta quando não é possível aplicar adequadamente as permissões necessárias
`AWS Lake Formation`	• Deriva e retorna corretamente as permissões efetivas para uma determinada entidade principal. • Valida os recursos suportados por terceiros em uma operação de API chamada por chamada. • Retorna credenciais do IAM com escopo reduzido somente quando os recursos anunciados do mecanismo correspondem aos definidos nos recursos do catálogo, caso contrário, retorna um erro.

Lake Formation fluxo de trabalho para operações de API de integração de aplicativos

A seguir está o fluxo de trabalho das operações da API de integração de aplicativos:

1. Um usuário envia uma consulta ou solicitação de dados usando um mecanismo de consulta integrado de terceiros. O mecanismo de consulta assume um perfil do IAM que representa o usuário ou um grupo de usuários e recupera credenciais confiáveis para serem usadas ao chamar as operações da API de integração de aplicativos.

2. O mecanismo de consulta chama `GetUnfilteredTableMetadata` e, se for uma tabela particionada, o mecanismo de consulta chama `GetUnfilteredPartitionsMetadata` para recuperar metadados e informações de política do catálogo de dados.
3. O Lake Formation realiza a autorização para a solicitação. Se o usuário não tiver as permissões apropriadas na mesa, ele será AccessDeniedExceptiondescartado.
4. Como parte da solicitação, o mecanismo de consulta envia a filtragem compatível. Há dois sinalizadores que podem ser enviados em uma matriz: COLUMN_PERMISSIONS e CELL_FILTER_PERMISSION. Se o mecanismo de consulta não oferecer suporte a nenhum desses recursos e existir uma política na tabela para o recurso, a será lançado e a consulta falhará. PermissionTypeMismatchException Isso acontece para evitar o vazamento de dados.
5. A resposta obtida contém o seguinte:
   - O esquema inteiro da tabela para que os mecanismos de consulta possam usá-lo para analisar os dados do armazenamento.
   - Uma lista de colunas autorizadas que o usuário tem acesso. Se a lista de colunas autorizadas estiver vazia, isso indica que o usuário tem permissões `DESCRIBE`, mas não tem permissões `SELECT`, e a consulta falha.
   - Um alerta, `IsRegisteredWithLakeFormation`, que indica se o Lake Formation pode fornecer credenciais para esses dados de recursos. Se isso retornar falso, as credenciais dos clientes devem ser usadas para acessar o Amazon S3.
   - Uma lista de `CellFilters`, se houver, que deve ser aplicada às linhas de dados. Essa lista contém colunas e uma expressão para avaliar cada linha. Isso só deve ser preenchido se CELL_FILTER_PERMISSION for enviado como parte da solicitação e houver um filtro de dados na tabela do usuário chamador.
6. Depois que os metadados são recuperados, o mecanismo de consulta chama `GetTemporaryGlueTableCredentials` ou obtém AWS credenciais `GetTemporaryGluePartitionCredentials` para recuperar dados da localização do Amazon S3.
7. O mecanismo de consulta lê objetos relevantes do Amazon S3, filtra os dados com base nas políticas recebidas na etapa 2 e retorna os resultados ao usuário.

As operações da API de integração de aplicativos para o Lake Formation contêm conteúdo adicional para configurar a integração com mecanismos de consulta de terceiros. Você pode ver os informações da operação na seção Operações da API de fornecimento de credenciais.

`QuerySessionContext`É uma estrutura que os mecanismos de consulta também podem enviar Lake Formation para essas operações de API de integração de aplicativos. Ele permite Lake Formation armazenar e utilizar contexto adicional para uma determinada consulta. Veja a seguir um exemplo de como QuerySessionContextdeve ser usado:

1. O mecanismo de consulta faz uma `GetInternalUnfilteredMetadata` chamada, passando uma estrutura QSC contendo um ID de consulta exclusivo na solicitação:

```
{
    "QuerySessionContext": {
        "QueryId": "your-unique-identifier-here"
    }
}
```

2. A `GetInternalUnfilteredMetadata` chamada retornará uma `QueryAuthorizationId` string na resposta. Na próxima chamada de consulta (e em qualquer chamada subsequente) que aceita uma estrutura QSC na entrada, o mecanismo de consulta passa a mesma estrutura QSC que agora também contém o `QueryAuthorizationId` retornado por. Lake Formation Suponha que a próxima chamada seja`GetTemporaryGlueTableCredentials`; a solicitação conterá:

```
{
    "QuerySessionContext": {
        "QueryAuthorizationId": "lf-returned-query-authz-id-here",
        "QueryId": "your-unique-identifier-here"
    },
}
```

Como registrar um mecanismo de consulta de terceiros

Antes que um mecanismo de consulta de terceiros possa usar as operações da API de integração de aplicativos, você precisa habilitar explicitamente as permissões para que o mecanismo de consulta chame as operações da API em seu nome. Isso é feito em algumas etapas:

1. Você precisa especificar as AWS contas e as tags de sessão do IAM que exigem permissão para chamar as operações da API de integração de aplicativos por meio do AWS Lake Formation console, do AWS CLI ou do API/SDK.
2. Quando o mecanismo de consulta de terceiros assume o perfil de execução em sua conta, o mecanismo de consulta deve anexar uma tag de sessão registrada no Lake Formation

representando o mecanismo de terceiros. O Lake Formation usa essa tag para validar se a solicitação for proveniente de um mecanismo aprovado. Para obter mais informações sobre tags de sessão, consulte Tags de sessão no Guia do usuário do IAM.

3. Ao configurar um perfil de execução de mecanismo de consulta de terceiros, você deve ter o seguinte conjunto mínimo de permissões na política do IAM:

   JSON

```
{
  "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. Configure uma política de confiança de perfil no perfil de execução do mecanismo de consulta para ter um controle de acesso preciso sobre qual par de chave-valor de tag de sessão pode ser anexado a esse perfil. No exemplo a seguir, esse perfil só pode ter a chave `"LakeFormationAuthorizedCaller"` da tag da sessão e o valor `"engine1"` da tag da sessão a serem anexados, e nenhum outro par de chave e valor da tag da sessão é permitido.

```
{
    "Sid": "AllowPassSessionTags",
    "Effect": "Allow",
    "Principal": {
        "AWS": "arn:aws:iam::111122223333:role/query-execution-role"
    },
    "Action": "sts:TagSession",
    "Condition": {
    "StringLike": {
```

```
            "aws:RequestTag/LakeFormationAuthorizedCaller": "engine1"          }
        }
}
```

Quando `LakeFormationAuthorizedCaller` chama a operação da STS:AssumeRole API para buscar credenciais para o mecanismo de consulta usar, a tag da sessão deve ser incluída na AssumeRole solicitação. A credencial temporária retornada pode ser usada para fazer solicitações de API de integração de aplicativos do Lake Formation.

As operações da API de integração de aplicativos do Lake Formation exigem que a entidade principal da chamada tenha um perfil do IAM. O perfil do IAM deve incluir uma tag de sessão com um valor predeterminado que tenha sido registrado com o Lake Formation. Essa tag permite ao Lake Formation verificar se o perfil usada para chamar as operações da API de integração de aplicativos tem permissão para fazer isso.

Como habilitar permissões para que um mecanismo de consulta de terceiros chame operações de API de integração de aplicativos

Siga estas etapas para permitir que um mecanismo de consulta terceirizado chame as operações da API de integração de aplicativos por meio do AWS Lake Formation console, AWS CLI ou API/SDK.

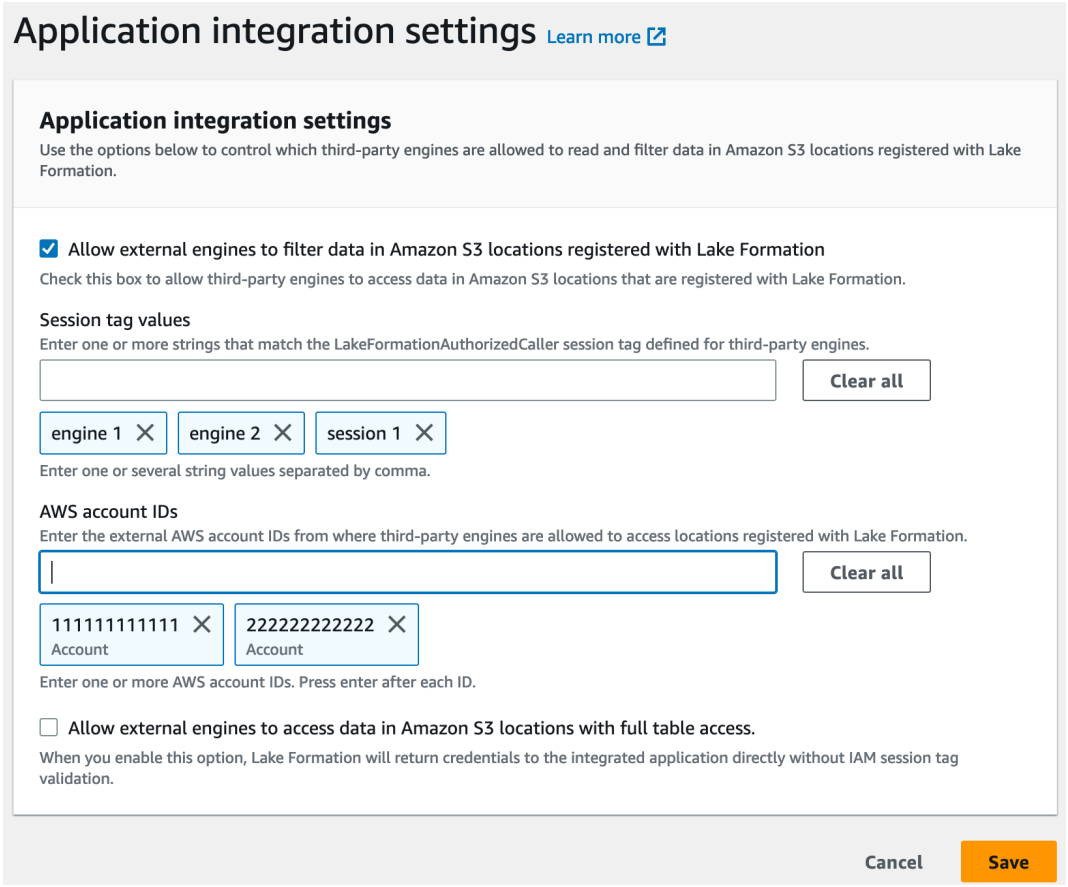
Console

Como registrar sua conta para filtragem externa de dados:

1. Faça login no Console de gerenciamento da AWS, e abra o console do Lake Formation em https://console.aws.amazon.com/lakeformation/.
2. No painel de navegação à esquerda, expanda Administração e selecione Configurações da integração de aplicações.
3. Na página de configuração de Integração do aplicativo, selecione a opção Permitir que mecanismos externos filtrem dados em locais do Amazon S3 registrados com o Lake Formation.
4. Digite as tags de sessão que você criou para o mecanismo de terceiros. Para obter informações sobre tags de sessão, consulte Passando tags de sessão no AWS STS no Guia AWS Identity and Access Management do usuário.

5. Digite os IDs de conta dos usuários que podem usar o mecanismo de terceiros para acessar informações de metadados não filtradas e as credenciais de acesso aos dados dos recursos na conta atual.

   Você também pode usar o campo ID da AWS conta para configurar o acesso entre contas.



CLI

Use o comando `put-data-lake-settings` do CLI para definir os parâmetros a seguir.

Há três campos a serem configurados ao usar esse AWS CLI comando:

- `allow-external-data-filtering` — (boolean) Indica que um mecanismo de terceiros pode acessar informações de metadados não filtradas e credenciais de acesso a dados de recursos na conta corrente.
- `external-data-filtering-allow-list` — (matriz) Uma lista de IDs de conta que podem acessar informações de metadados não filtradas e credenciais de acesso a dados de recursos na conta atual ao usar um mecanismo de terceiros. Quando AllowExternalDataFiltering definido

como verdadeiro, a ExternalDataFilteringAllowList propriedade deve incluir pelo menos um ID de conta. Não é permitida uma lista vazia.

- `authorized-sessions-tag-value-list` — (matriz) Uma lista de valores de tags de sessão autorizados (cadeias). Se uma credencial de perfil do IAM tiver sido anexada a um par de chave e valor autorizado, se a tag da sessão for incluída na lista, a sessão terá acesso a informações de metadados não filtrados e credenciais de acesso a dados em recursos na conta configurada. A chave da tag de sessão autorizada é definida como `*LakeFormationAuthorizedCaller*`.
- `AllowFullTableExternalDataAccess` - (booleano) Permitir ou não que um mecanismo de consulta de terceiros obtenha credenciais de acesso a dados sem tags de sessão quando um chamador tiver permissões completas de acesso a dados.

Por exemplo:

```
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

Use a operação `PutDataLakeSetting` da API para definir os seguintes parâmetros.

Há três campos a serem configurados ao usar essa operação de API:

- `AllowExternalDataFiltering` — (boolean) Indica se um mecanismo de terceiros pode acessar informações de metadados não filtradas e credenciais de acesso a dados de recursos nesta conta.
- `ExternalDataFilteringAllowList` — (matriz) Uma lista de IDs de conta que podem acessar informações de metadados não filtradas e as credenciais de acesso aos dados dos recursos na conta atual usando um mecanismo de terceiros.
- `AuthorizedSectionsTagValueList` — (matriz) Uma lista de valores de tag autorizados (cadeias). Se uma credencial de perfil do IAM tiver sido anexada a uma tag autorizada, a sessão terá acesso às informações de metadados não filtradas e às credenciais de acesso aos dados nos recursos da conta configurada. A chave da tag de sessão autorizada é definida como `*LakeFormationAuthorizedCaller*`.
- `AllowFullTableExternalDataAccess` - (booleano) Permitir ou não que um mecanismo de consulta de terceiros obtenha credenciais de acesso a dados sem tags de sessão quando um chamador tiver permissões completas de acesso a dados.

Por exemplo:

```
//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));
}
```

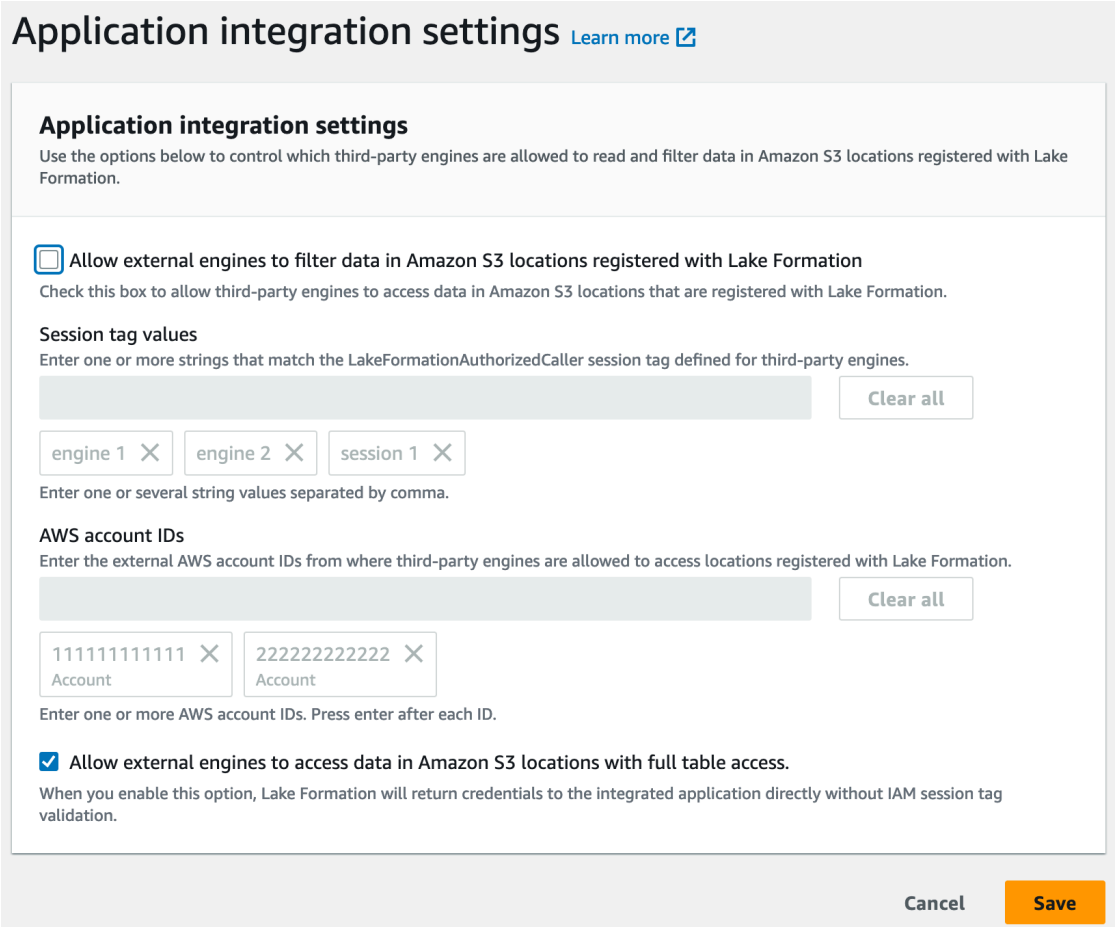
Integração de aplicativos para acesso total à tabela

Siga estas etapas para permitir que mecanismos de consulta de terceiros acessem dados sem a validação da tag de sessão do IAM:

Console

1. Faça login no console do Lake Formation em https://console.aws.amazon.com/lakeformation/.
2. No painel de navegação à esquerda, expanda Administração e selecione Configurações da integração de aplicações.
3. Na página Configurações da integração de aplicações, marque a caixa de seleção Permitir que mecanismos externos acessem dados em locais do Amazon S3 com acesso total à tabela.

   Quando você habilita essa opção, o Lake Formation exibe as credenciais diretamente para a aplicação de consulta, sem a validação da tag de sessão do IAM.



AWS CLI

Use o comando `put-data-lake-settings` do CLI para definir os parâmetros em `AllowFullTableExternalDataAccess`.

```
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
    }
}
```

Trabalhando com outros AWS serviços

AWS serviços como Amazon Athena AWS Glue, Amazon Redshift Spectrum e Amazon EMR podem ser AWS Lake Formation usados para acessar com segurança dados em locais do Amazon S3 registrados no Lake Formation. Com o Lake Formation, você pode definir e gerenciar permissões refinadas de controle de acesso (FGAC) para suas tabelas no AWS Glue Data Catalog. Cada um desses AWS serviços é um chamador confiável para o Lake Formation, e o Lake Formation fornece acesso aos dados armazenados no Amazon S3 por meio de credenciais temporárias. Para obter mais informações, consulte Como funciona a integração de aplicações do Lake Formation.

Para aproveitar esses recursos, o Lake Formation exige que você primeiro registre a localização do Amazon S3 e atribua as permissões apropriadas à entidade principal do IAM para acessar a tabela, o banco de dados e a localização do Amazon S3. Para obter mais informações, consulte, Gerenciando permissões do Lake Formation.

As tabelas a seguir listam os tipos de permissões do Lake Formation suportadas pelo Amazon Athena, AWS Glue Amazon EMR e Amazon Redshift Spectrum para acessar dados AWS Glue de tabelas padrão e tabelas transacionais (Apache Iceberg, Apache Hudi e Linux Foundation Delta Lake) com dados armazenados no Amazon S3 e metadados de tabelas no Catálogo de dados.

AWS serviços e tipos de permissão compatíveis para tabelas e exibições AWS Glue padrão

AWS serviço	Permissões no nível de tabela	Permissões no nível de coluna	Permissões no nível de linha e célula
Athena SQL	Acesso de leitura/gravação	Acesso de leitura	Acesso de leitura
Athena Spark	Sem compatibilidade	Sem compatibilidade	Sem compatibilidade
Redshift Spectrum em um cluster provisionado ou Amazon Redshift sem servidor	Acesso de leitura/gravação	Acesso de leitura	Acesso de leitura
Apache Spark no Amazon EMR (EC2)	Acesso de leitura/gravação	Acesso de leitura	Acesso de leitura

AWS serviço	Permissões no nível de tabela	Permissões no nível de coluna	Permissões no nível de linha e célula
Apache Hive no Amazon EMR (EC2)	Acesso de leitura/gravação	Acesso de leitura	Não compatível
Apache Spark no EMR Sem Servidor	Acesso de leitura/gravação	Acesso de leitura	Acesso de leitura
Apache Hive no EMR Sem Servidor	Sem compatibilidade	Sem compatibilidade	Sem compatibilidade
Amazon EMR no EKS	Sem compatibilidade	Sem compatibilidade	Sem compatibilidade
AWS Glue ETL	Acesso de leitura/gravação	AWS Glue 5.0 ou superior oferece suporte ao acesso de leitura.	AWS Glue 5.0 ou superior oferece suporte ao acesso de leitura.

Considerações e limitações

- O Athena Spark não permite consultas a tabelas do Catálogo de Dados com permissões do Lake Formation.
- Os usuários baseados em SAML do Athena podem ler fontes de dados protegidas usando as permissões do Lake Formation ao habilitarem a federação baseada em SAML 2.0. Os usuários do SAML podem inserir dados nas tabelas do Parquet.
- O Apache Spark no EMR Sem Servidor não permite consultas a visualizações do Catálogo de Dados.
- O Apache Hive no EMR Sem Servidor não permite consultas a tabelas com permissões do Lake Formation.
- AWS Glue 5.0 ou superior oferece suporte a controles de acesso refinados nas tabelas Iceberg e Hive no Catálogo de Dados que são apoiados pelo S3. Esse recurso permite que você configure controles de acesso em nível de tabela, linha, coluna e célula para consultas de leitura em suas AWS Glue tarefas do Apache Spark.

  Para obter mais informações, consulte versões do AWS Glue.

AWS serviços e tipos de permissão compatíveis para formatos de tabela transacional

AWS serviço	Iceberg	Hudi	Delta Lake (nativo)	Delta Lake (tabelas symlink)
Athena SQL	Compatível com leitura de tabelas com permissões no nível de tabela, coluna, linha e célula. Operações de gravação exigem acesso total à tabela.	Compatível com operações de leitura e criação de tabelas com permissões no nível de tabela, coluna, linha e célula. Operações de gravação não são compatíveis.	O Athena (versão 3 do mecanismo) permite leitura de tabelas nativas do Delta Lake com permissões no nível de tabela, coluna, linha e célula. Operações de gravação não são compatíveis.	O Athena (versão 3 do mecanismo) permite leitura de tabelas symlink do Delta Lake com permissões no nível de tabela, coluna, linha e célula. Operações de gravação não são compatíveis.
Redshift Spectrum em um cluster provisionado	Compatível com leitura de tabelas com permissões no nível de tabela, coluna, linha e célula. Operações de gravação não são compatíveis.	Compatível com leitura de tabelas com permissões no nível de tabela, coluna, linha e célula. Operações de gravação não são compatíveis.	Não compatível	Permite leitura de tabelas do Delta Lake por meio de manifesto symlink com permissões no nível de tabela, coluna, linha e célula. Operações de gravação não são compatíveis.
Apache Spark no Amazon EMR (EC2)	Compatível com leitura de tabelas com permissões no nível de	Compatível com leitura de tabelas com permissões no nível de	Compatível com leitura de tabelas com permissões no nível de	Compatível com leitura de tabelas com permissões no nível de

AWS serviço	Iceberg	Hudi	Delta Lake (nativo)	Delta Lake (tabelas symlink)
	tabela, coluna, linha e célula. Operações de gravação exigem acesso total à tabela.	tabela, coluna, linha e célula. Operações de gravação exigem acesso total à tabela.	tabela, coluna, linha e célula. Operações de gravação não são compatíveis.	tabela, coluna, linha e célula. Operações de gravação exigem acesso total à tabela.
AWS Glue ETL	AWS Glue 5.0 ou superior suporta a leitura de tabelas com permissões em nível de tabela, coluna, linha e célula.	Suporte read/ write em tabelas com permissões em nível de tabela.	Suporte read/ write em tabelas com permissões em nível de tabela.	Suporte read/ write em tabelas com permissões em nível de tabela.

Tópicos

- Usando AWS Lake Formation com o Amazon Athena
- Usando AWS Lake Formation com o Amazon Redshift Spectrum
- Usando AWS Lake Formation com AWS Glue
- Usando AWS Lake Formation com o Amazon EMR
- Usando AWS Lake Formation com o Quick
- Usando AWS Lake Formation com o AWS CloudTrail Lake

Usando AWS Lake Formation com o Amazon Athena

Amazon Athena é um serviço de consulta sem servidor que ajuda a analisar dados estruturados, semiestruturados e não estruturados armazenados no Amazon S3. Você pode usar o Athena SQL para consultar dados dos formatos de dados CSV, JSON, Parquet e Avro. O Athena SQL também é compatível com formatos de tabela como Apache Hive, Apache Hudi e Apache Iceberg. O Athena se integra ao AWS Glue Data Catalog para armazenar metadados de seus conjuntos de dados no

Amazon S3. O Athena pode usar o Lake Formation para definir e manter políticas de controle de acesso nesses conjuntos de dados.

Aqui estão alguns casos de uso comuns em que você pode usar o Lake Formation com o Athena.

- Use as permissões do Lake Formation para acessar os recursos do catálogo de dados (banco de dados e tabelas) do Athena. Você pode usar o método de recurso nomeado ou as tags do LF para definir permissões no banco de dados e nas tabelas. Para obter mais informações, consulte:
  - Conceder permissões de banco de dados usando o método de recurso nomeado
  - Controle de acesso baseado em tags do Lake Formation

  > **Note**
  >
  > As permissões do Lake Formation são válidas somente ao usar o Athena SQL para consultar dados de origem do Amazon S3 e metadados no Catálogo de Dados.
  > O Athena Spark não permite consultas a tabelas do Catálogo de Dados com permissões do Lake Formation. As permissões do Lake Formation oferecem suporte a operações de leitura e gravação em bancos de dados e tabelas.

  > **Note**
  >
  > Você não pode aplicar filtros de dados ao usar tags do LF para gerenciar permissões nos recursos do catálogo de dados.

- Controle os resultados da consulta usando a Filtros de dados no Lake Formation para proteger tabelas em seus data lakes do Amazon S3, concedendo permissões nos níveis de coluna, linha e célula. Veja a limitação na projeção de partições no Guia do usuário do Amazon Athena.
- Aplique um controle de acesso refinado dos dados disponíveis para o usuário do Athena baseado em SAML ao executar consultas federadas.

  Os drivers Athena JDBC e ODBC oferecem suporte à configuração do acesso federado à sua fonte de dados usando o provedor de identidades (IdP) baseado em SAML. Use o Quick integrado ao Lake Formation com sua função existente do IAM ou usuários ou grupos do SAML para visualizar os resultados da consulta do Athena.

> **Note**
>
> As permissões do Lake Formation para usuários e grupos SAML serão aplicadas somente quando você enviar consultas ao Athena usando o driver JDBC ou ODBC.

Para obter mais informações, consulte Como usar o Lake Formation e drivers JDBC e ODBC do Athena para acesso federado ao Athena.

> **Note**
>
> Atualmente, não há suporte para autorizar o acesso às identidades SAML no Lake Formation nas seguintes regiões:
>
> - Oriente Médio (Bahrein): me-south-1
> - Ásia-Pacífico (Hong Kong): ap-east-1
> - África (Cidade do Cabo): af-south-1
> - China (Ningxia): cn-northwest-1
> - Asia Pacific (Osaka): ap-northeast-3

- Use Cross-account compartilhamento de dados em Lake Formation para consultar tabelas em outra conta.

> **Note**
>
> Para obter mais informações sobre limitações ao usar as permissões do Lake Formation para `Views`, consulte Considerações e limitações.

Suporte a formatos de tabelas transacionais

A aplicação das permissões do Lake Formation permite que você proteja seus dados transacionais em seus data lakes baseados no Amazon S3. A tabela abaixo lista os formatos de tabela transacional compatíveis com as permissões do Athena e do Lake Formation. O Lake Formation impõe essas permissões quando os usuários do Athena executam suas consultas.

Formato da tabela	Descrição e operações permitidas	Permissões do Lake Formation possíveis no Athena
Apache Hudi	Um formato usado para simplificar o processamento incremental de dados e o desenvolvimento de pipelines de dados. O Athena oferece suporte a operações de criação e leitura usando formatos de tabela Apache Hudi em conjuntos de dados do Amazon S3 para os tipos de tabela Hudi Copiar na Gravação (CoW) e Mesclar na Leitura (MoR). O Athena não permite operações de gravação em tabelas Hudi. Use o Athena para consultar conjuntos de dados Hudi.	Use Filtragem de dados e segurança por célula no Lake Formation para proteger a tabela Hudi com permissões no nível de tabela, coluna, linha e célula.
Apache Iceberg	Um formato de tabela aberto que gerencia grandes coleções de arquivos como tabelas e oferece suporte a operações analíticas modernas de data lake, como inserção, atualização, exclusão e consultas de viagem no tempo em nível de registro. Para obter mais informações sobre o suporte do Athena	Suporte para permissões em nível de tabela, coluna, linha e célula. Atualmente, o Lake Formation não oferece suporte ao gerenciamento de permissões em operações de gravação como `VACUUM`, `MERGE`, `UPDATE` e `OPTIMIZE` em tabelas em formatos de tabela aberta.

Formato da tabela	Descrição e operações permitidas	Permissões do Lake Formation possíveis no Athena
	para tabelas Iceberg, consulte Como usar tabelas Iceberg.	
Linux Foundation Delta Lake	O Delta Lake é um projeto de código aberto que ajuda a implementar arquitetu ras modernas de data lake, geralmente construídas no Amazon S3 ou no Sistema de Arquivos Distribuído do Hadoop (HDFS). O Athena é compatível com tabelas Delta Lake criadas usando uma definição de tabela de manifesto baseada em links simbólicos a AWS Glue Data Catalog partir de uma tabela Delta Lake. Para obter mais informações, consulte Rastrear tabelas do Delta Lake usando AWS Glue rastreadores. O Athena (motor de versão 3) suporta a leitura de tabelas nativas do Delta Lake. Para obter mais informaçõ es, consulte Apresentando o suporte de mesa nativo do Delta Lake com AWS Glue rastreadores.	Suporte para permissões em nível de tabela, coluna, linha e célula para tabelas de links simbólicos e tabelas nativas do Delta Lake.

Recursos adicionais do

Publicações em blogs, vídeos e oficinas

- Como consultar um conjunto de dados do Apache Hudi em um data lake do Amazon S3 com o Amazon Athena
- Crie um data lake Apache Iceberg usando o Amazon Athena, o Amazon EMR e AWS Glue
- Insira, atualize e exclua no Amazon S3 com Athena e Apache Iceberg
- Oficina de Controle de acesso baseado em tag do LF do Lake Formation sobre como consultar um data lake.

Usando AWS Lake Formation com o Amazon Redshift Spectrum

O Amazon Redshift Spectrum permite que você consulte e recupere dados em data lakes do Amazon S3 sem que seja necessário carregar dados em nós de cluster do Amazon Redshift.

O Redshift Spectrum oferece suporte a duas formas de registrar um catálogo de AWS Glue dados externo habilitado com o Lake Formation.

- Como usar um perfil do IAM anexado ao cluster que tenha permissão para acessar o catálogo de dados

  Para criar um perfil do IAM, siga as etapas descritas no procedimento abaixo.

  Controlando o acesso ao AWS Glue Data Catalog
- Como usar identidade federada do IAM configurada para gerenciar o acesso a recursos AWS Glue Data Catalog externos

  O Redshift Spectrum suporta a consulta de tabelas do Lake Formation usando identidades federadas do IAM. As identidades do IAM podem ser um usuário do IAM ou um perfil do IAM. Para obter mais informações sobre a federação de identidades IAM no Redshift Spectrum, consulte Como usar uma identidade federada para gerenciar o acesso do Amazon Redshift a recursos locais e tabelas externas do Redshift Spectrum.

Com a integração do Lake Formation com o Redshift Spectrum, você pode definir permissões de controle de acesso em nível de linha, coluna e célula nas tabelas depois que seus dados forem registrados no Lake Formation.

Para obter mais informações, consulte Usando o Redshift Spectrum com. AWS Lake Formation

O Redshift Spectrum suporta leituras ou consultas `SELECT` nas tabelas de esquema externo gerenciadas pelo Lake Formation.

Para obter mais informações, confira Como criar esquemas externos para Redshift Spectrum.

Suporte para tipos de tabelas transacionais

Esta tabela lista os formatos de tabela transacional suportados no Redshift Spectrum e as permissões aplicáveis do Lake Formation.

Formatos de tabela compatíveis

Formato da tabela	Descrição e operações permitidas	Permissões do Lake Formation compatíveis com o Redshift Spectrum
Apache Hudi	Um formato usado para simplificar o processamento incremental de dados e o desenvolvimento de pipelines de dados. O Redshift Spectrum suporta operações de gravação de inserção, exclusão e inserção usando o formato de tabela Apache Hudi Copiar na Gravação (CoW) no Amazon S3. Para obter mais informaçõ es, consulte Criação de tabelas externas para dados gerenciados no Apache Hudi.	Use Filtragem de dados e segurança por célula no Lake Formation para proteger a tabela Hudi com permissõe s no nível de tabela, coluna, linha e célula.
Apache Iceberg	Um formato de tabela aberto que gerencia grandes coleções de arquivos como	O Redshift Spectrum oferece suporte a tabelas do Apache Iceberg para consultas.

Formato da tabela	Descrição e operações permitidas	Permissões do Lake Formation compatíveis com o Redshift Spectrum
	tabelas e oferece suporte a operações analíticas modernas de data lake, como inserção, atualizaç ão, exclusão e consultas de viagem no tempo em nível de registro. Para obter mais informaçõ es, confira Como usar tabelas do Apache Iceberg com o Amazon Redshift.	
Linux Foundation Delta Lake	O Delta Lake é um projeto de código aberto que ajuda a implementar arquitetu ras modernas de data lake, geralmente construídas no Amazon S3 ou no Sistema de Arquivos Distribuído do Hadoop (HDFS). O Redshift Spectrum oferece suporte à consulta de tabelas Delta Lake. Para obter mais informações, consulte Criação de tabelas externas para dados gerenciados no Delta Lake.	Suporte para permissões em nível de tabela, coluna, linha e célula.

Recursos adicionais do

Publicações em blogs e oficinas

- Centralize a governança do seu data lake usando, AWS Lake Formation ao mesmo tempo, uma arquitetura de dados moderna com o Amazon Redshift Spectrum
- Use o Redshift Spectrum para consultar as tabelas do Apache HUDI Copiar na Gravação (CoW) no data lake do Amazon S3

Usando AWS Lake Formation com AWS Glue

Engenheiros e DevOps profissionais de dados usam AWS Glue o Extract, Transform and Load (ETL) com o Apache Spark para realizar transformações em seus conjuntos de dados no Amazon S3 e carregar os dados transformados em lagos de dados e armazéns de dados para análise, aprendizado de máquina e desenvolvimento de aplicativos. Com equipes diferentes acessando o mesmo conjunto de dados no Amazon S3, é imperativo conceder e restringir permissões com base em seus perfis.

AWS Lake Formation é construído e AWS Glue os serviços interagem das seguintes maneiras:

- Lake Formation e AWS Glue compartilham o mesmo catálogo de dados.
- Os seguintes atributos do console Lake Formation invocam o console AWS Glue:
  - Trabalhos — Para obter mais informações, consulte Como adicionar trabalhos no Guia do desenvolvedor do AWS Glue .
  - Crawlers – Para obter mais informações, consulte Catalogação de tabelas com um Crawler Guia do desenvolvedor do AWS Glue .
- Os fluxos de trabalho gerados quando você usa um esquema do Lake Formation são fluxos de trabalho AWS Glue. Você pode visualizar e gerenciar esses fluxos de trabalho no console do Lake Formation e no console AWS Glue.
- As transformações de machine learning são fornecidas com o Lake Formation e são baseadas em operações de API do AWS Glue. Você cria e gerencia transformações de machine learning no console AWS Glue. Para obter mais informações, consulte Transformações de machine learning no Guia do desenvolvedor do AWS Glue .

Você pode usar o controle de acesso refinado do Lake Formation para gerenciar seus recursos existentes do catálogo de dados e os locais de dados do Amazon S3.

> **Note**
>
> AWS Glue 5.0 ou superior oferece suporte a controles de acesso refinados nas tabelas Iceberg e Hive que são apoiadas pelo S3. Esse recurso permite que você configure controles de acesso em nível de tabela, linha, coluna e célula para consultas de leitura em suas AWS Glue tarefas do Apache Spark.

Suporte para tipos de tabelas transacionais

A aplicação das permissões do Lake Formation permite que você proteja seus dados transacionais em seus data lakes baseados no Amazon S3. A tabela abaixo lista os formatos de tabela transacional suportados AWS Glue e as permissões do Lake Formation. Lake Formation impõe essas permissões para AWS Glue operações.

Formatos de tabela compatíveis

Formato da tabela	Descrição e operações permitidas	Permissões do Lake Formation suportadas em AWS Glue
Apache Hudi	Um formato de tabela aberta usado para simplificar o processamento incremental de dados e o desenvolvimento de pipelines de dados. Para exemplos, consulte Usando a estrutura Hudi em AWS Glue.	As permissões em nível de tabela estão disponíveis para tabelas do Hudi. Para obter mais informações, consulte Limitações.
Apache Iceberg	Um formato de tabela aberta que gerencia grandes coleções de arquivos como tabelas. Para obter exemplos, consulte Usando a estrutura Iceberg em AWS Glue.	AWS Glue a versão 5.0 e superior permite configura r controles de acesso em nível de tabela, linha, coluna e célula para consultas de leitura em suas tarefas do

Formato da tabela	Descrição e operações permitidas	Permissões do Lake Formation suportadas em AWS Glue
		Apache Spark AWS Glue para tabelas Iceberg. Para obter mais informações, consulte Limitações.
Linux Foundation Delta Lake	O Delta Lake é um projeto de código aberto que ajuda a implementar arquitetu ras modernas de data lake, geralmente construídas no Amazon S3 ou no Sistema de Arquivos Distribuído do Hadoop (HDFS). Para ver exemplos, consulte Usando a estrutura Delta Lake em AWS Glue.	As permissões em nível de tabela estão disponíveis para tabelas do Delta Lake. Para obter mais informações, consulte Limitações.

Recursos adicionais do

Publicações em blogs e repositórios

- Use o AWS Glue conector para ler e gravar tabelas Apache Iceberg com transações ACID e realizar viagens no tempo
- Escrevendo em tabelas do Apache Hudi usando conector personalizado AWS Glue
- AWS repositório do modelo Cloudformation e amostra de código do pyspark para analisar dados de streaming usando o Apache Hudi e o AWS Glue Amazon S3.

Usando AWS Lake Formation com o Amazon EMR

O Amazon EMR é uma plataforma de cluster AWS gerenciada flexível na qual você pode executar qualquer código personalizado em estruturas de big data compatíveis, como Hadoop Map-Reduce,

Spark, Hive, Presto etc. As organizações também usam o Amazon EMR para executar aplicativos de processamento de dados em lote e streaming em um cluster altamente distribuído. Ao usar o Apache Spark no Amazon EMR, você pode executar transformações de dados e código personalizado em bancos de dados e tabelas cujas permissões são gerenciadas pelo Lake Formation.

Existem três opções para implementar o Amazon EMR:

- EMR no EC2
- Tecnologia sem servidor do EMR
- Amazon EMR no EKS

Para obter mais informações, consulte Integrar o Amazon EMR com o Lake Formation ou Usar o EMR Serverless com para um controle de acesso refinado AWS Lake Formation

Suporte a formatos de tabelas transacionais

As versões 6.15.0 e posteriores do Amazon EMR incluem suporte para permissões de controle de acesso em nível de tabela, linha, coluna e célula do Lake Formation nos formatos Apache Hudi , Apache Iceberg e Delta Lake ao ler e gravar dados com o Spark SQL.

Sobre limitações, consulte Considerations for Amazon EMR with Lake Formation.

Formatos de tabela compatíveis

Formato da tabela	Descrição e operações permitidas	Permissões do Lake Formation aceitas no Amazon EMR
Apache Hudi	Um formato de tabela aberta usado para simplificar o processamento incremental de dados e o desenvolvimento de pipelines de dados. Para obter uma lista das operações compatíveis, consulte Apache Hudi e Lake Formation.	O Amazon EMR oferece suporte ao controle de acesso no nível de tabela, linha, coluna e célula com o Apache Hudi.

Formato da tabela	Descrição e operações permitidas	Permissões do Lake Formation aceitas no Amazon EMR
Apache Iceberg	Um formato de tabela aberta que gerencia grandes coleções de arquivos como tabelas. Para obter uma lista das operações compatíveis, consulte Apache Iceberg e Lake Formation.	O Amazon EMR oferece suporte ao controle de acesso no nível de tabela, linha, coluna e célula com o Apache Iceberg.
Linux Foundation Delta Lake	O Delta Lake é um projeto de código aberto que ajuda a implementar arquitetu ras modernas de data lake, geralmente construídas no Amazon S3 ou no Sistema de Arquivos Distribuído do Hadoop (HDFS). Para obter uma lista das operações compatíveis, consulte Delta Lake e Lake Formation.	O Amazon EMR permite o controle de acesso no nível de tabela, linha, coluna e célula com tabelas do Delta Lake.

Recursos adicionais do

Guia do usuário, postagens de blog e oficinas

- Integração com o Amazon EMR usando perfis de runtime
- Comece a usar Apache Hudi, Apache Iceberg e Delta Lake com o Amazon EMR no EKS
- Usar o Delta Lake OSS com o EMR Sem Servidor

Usando AWS Lake Formation com o Quick

O Quick oferece suporte à exploração de conjuntos de dados gerenciados pelas permissões do Lake Formation no Amazon S3 usando o Athena.

Tanto os usuários das edições Standard quanto Enterprise do Quick se integram ao Lake Formation, mas de forma um pouco diferente.

- Edição corporativa — conceda permissões refinadas de controle de acesso (FGAC) a usuários e grupos individuais do Quick para acessar bancos de dados e tabelas.
- Edição padrão — conceda permissões às perfis do IAM para acessar bancos de dados e tabelas.

> **Note**
>
> Por padrão, o Quick usa uma função chamada`aws-quicksight-service-role-v0`. Você também pode definir funções personalizadas com as permissões necessárias que permitem que o Quick acesse o Athena.

Para obter mais informações, consulte Autorização de conexões por meio de AWS Lake Formation

Recursos adicionais do

Publicações no blog

- Habilite permissões refinadas para autores do Quick em AWS Lake Formation
- Analise seus dados com segurança com AWS Lake Formation e Quick

Usando AWS Lake Formation com o AWS CloudTrail Lake

AWS CloudTrail O Lake suporta a exploração de armazenamentos de dados Amazon Athena de eventos usando permissões refinadas em. AWS Lake Formation

> **Note**
>
> CloudTrail O lago só pode ser consultado. Amazon Athena

Para registrar seu armazenamento de dados de eventos do CloudTrail Lake no Lake Formation, consulte Federar um armazenamento de dados de eventos.

Registrando chamadas da API AWS Lake Formation usando AWS CloudTrail

AWS O Lake Formation é integrado com AWS CloudTrail, um serviço que fornece um registro das ações realizadas por um usuário, função ou AWS serviço no Lake Formation. CloudTrail captura todas as chamadas da API Lake Formation como eventos. As chamadas capturadas incluem chamadas do console do Lake Formation AWS Command Line Interface, do e chamadas de código para as ações da API do Lake Formation. Se você criar uma trilha, poderá habilitar a entrega contínua de CloudTrail eventos para um bucket do Amazon S3, incluindo eventos para Lake Formation. Se você não configurar uma trilha, ainda poderá ver os eventos mais recentes no CloudTrail console no Histórico de eventos. Usando as informações coletadas por CloudTrail, você pode determinar a solicitação que foi feita ao Lake Formation, o endereço IP do qual a solicitação foi feita, quem fez a solicitação, quando ela foi feita e detalhes adicionais.

Para saber mais sobre isso CloudTrail, consulte o Guia AWS CloudTrail do usuário.

Informações sobre Lake Formation em CloudTrail

CloudTrail é ativado por padrão quando você cria uma nova AWS conta. Quando a atividade ocorre no Lake Formation, essa atividade é registrada como um CloudTrail evento junto com outros eventos AWS de serviço no histórico de eventos. Um evento representa uma única solicitação de qualquer origem e inclui informações sobre a ação solicitada, a data e a hora da ação e os parâmetros de solicitação. Cada entrada de log ou evento contém informações sobre quem gerou a solicitação. As informações de identidade ajudam a determinar o seguinte:

- Se a solicitação foi feita com credenciais de usuário root ou AWS Identity and Access Management (IAM).
- Se a solicitação foi feita com credenciais de segurança temporárias de um perfil ou de um usuário federado.
- Se a solicitação foi feita por outro AWS serviço.

Para obter mais informações, consulte Elemento userIdentity do CloudTrail .

Você pode visualizar, pesquisar e baixar eventos recentes para sua AWS conta. Para obter mais informações, consulte Visualização de eventos com histórico de CloudTrail eventos.

Para um registro contínuo dos eventos em sua AWS conta, incluindo eventos do Lake Formation, crie uma trilha. Uma trilha permite CloudTrail entregar arquivos de log para um bucket do Amazon S3. Por padrão, quando você cria uma trilha no console, ela é aplicada a todas as regiões do AWS . A trilha registra eventos de todas as regiões na AWS partição e entrega os arquivos de log ao bucket do Amazon S3 que você especificar. Além disso, você pode configurar outros AWS serviços, como Amazon Athena, para analisar e agir com base nos dados de eventos coletados nos CloudTrail registros. CloudTrail também pode entregar arquivos de log para Amazon CloudWatch Logs and CloudWatch Events.

Para saber mais, consulte:

- Visão geral da criação de uma trilha
- CloudTrail serviços e integrações suportados
- Configurando notificações do Amazon SNS para CloudTrail
- Recebendo arquivos de CloudTrail log de várias regiões e Recebendo arquivos de CloudTrail log de várias contas

Noções básicas de eventos do Lake Formation

Todas as ações da API Lake Formation são registradas CloudTrail e documentadas no Guia do AWS Lake Formation Desenvolvedor. Por exemplo, chamadas para as `RevokePermissions` ações `PutDataLakeSettingsGrantPermissions`, e geram entradas nos arquivos de CloudTrail log.

O exemplo a seguir mostra um CloudTrail evento para a `GrantPermissions` ação. A entrada inclui o usuário que concedeu a permissão (`datalake_admin`), a entidade principal à qual a permissão foi concedida (`datalake_user1`) e a permissão que foi concedida (`CREATE_TABLE`). A entrada também mostra que a concessão falhou porque o banco de dados de destino não foi especificado no argumento `resource`.

```
{
    "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"
}
```

O próximo exemplo mostra uma entrada de CloudTrail registro para a `GetDataAccess` ação. As entidades principais não chamam essa API diretamente. Em vez disso, `GetDataAccess` é registrado sempre que um AWS serviço principal ou integrado solicita credenciais temporárias para acessar dados em um local de data lake registrado no 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"
    },
...
}
```

Rastreando o acesso aos dados CloudTrail com o Lake Formation

O Lake Formation emite credenciais temporárias com escopo reduzido para mecanismos de consulta, como Amazon EMR AWS Glue e, para acessar locais de dados de tabelas no Amazon S3. O Lake Formation usa a função do IAM que você especifica ao registrar um local de dados no Lake Formation para vender essas credenciais.

Mudança no comportamento da CloudTrail exploração madeireira (fevereiro de 2026)

Antes de fevereiro de 2026, quando um mecanismo de consulta fazia uma `s3:GetObject` chamada para buscar dados da consulta, os CloudTrail registros exibiam a função IAM do registro do Lake Formation como principal.

A partir de fevereiro de 2026, CloudTrail os registros de `s3:GetObject` chamadas também incluem o identificador exclusivo do IAM (começando comAROA) da função do IAM que iniciou a consulta. Esse identificador corresponde à função do IAM que iniciou a consulta, ao trabalho do Amazon EMR ou ao trabalho de AWS Glue ETL.

Rastreando o acesso aos dados desde o início da consulta até o Amazon S3

Cada `s3:GetObject` chamada recebida CloudTrail corresponde a uma `lakeformation:GetDataAccess` chamada. Você pode usar o identificador de função do IAM que está presente nas duas entradas de registro para rastrear a função do IAM do chamador desde o início da consulta até o acesso aos dados no Amazon S3.

Para habilitar as informações de identidade de origem nos CloudTrail registros de eventos de dados do Amazon S3, siga as etapas abaixo. Para habilitar eventos de dados do Amazon S3 CloudTrail, consulte Habilitar CloudTrail o registro de eventos para buckets e objetos do Amazon S3.

> **Important**
>
> A ativação de eventos de dados do Amazon S3 CloudTrail incorre em cobranças adicionais.

Para habilitar o rastreamento de identidade de origem nos CloudTrail registros

1. Usando, execute o `get-data-lake-settings` comando. Na resposta, edite o `Parameters` campo a ser adicionado`"SET_SOURCE_IDENTITY": "TRUE"`.

   Execute o comando `put-data-lake-settings`.

   ```
   "Parameters": {
       "CROSS_ACCOUNT_VERSION": "4",
       "SET_SOURCE_IDENTITY": "TRUE"
   },
   ```

2. Na política de confiança da função do IAM usada para registrar sua localização de dados do Amazon S3 no Lake Formation, adicione: `sts:SetSourceIdentity`

   ```
   {
       "Version": "2012-10-17",
       "Statement": [
           {
               "Effect": "Allow",
               "Principal": {
                   "Service": [
                       "lakeformation.amazonaws.com"
                   ]
               },
               "Action": [
                   "sts:AssumeRole",
                   "sts:SetSourceIdentity"
               ]
           }
       ]
   }
   ```

O registro a seguir mostra as informações de identidade de origem em um `s3:GetObject` CloudTrail registro de exemplo.

```
"sessionContext": {
    "sessionIssuer": {
        "type": "Role",
        "principalId": "AROA2EXAMPLEXYZ3AB6CDKLM",
        "arn": "arn:aws:iam::111122223333:role/LFRegistrationRole",
        "accountId": "111122223333",
        "userName": "LFRegistrationRole"
    },
    "sourceIdentity": "V2.111122223333.AROA3EXAMPLEPQR7ST9UVWXY"
},
"invokedBy": "glue.amazonaws.com"
```

Consulte também

- Para obter informações sobre como rastrear o contexto do usuário do IAM Identity Center em CloudTrail registros, consulteIntegrar o Centro de Identidade do IAM.

Monitoramento do acesso ao catálogo federado do Amazon Redshift

Os catálogos federados do Amazon Redshift no não geram eventos nos AWS Glue Data Catalog registros do `GetDataAccess` Lake Formation. CloudTrail Para acompanhar às tabelas do Amazon Redshift no Data Catalog, monitore eventos `GetTable` e `BatchGetTable`, que capturam padrões de acesso a metadados para tabelas federadas do Redshift, incluindo nome da tabela, nome do banco de dados e informações de ID do catálogo.

Consulte também

- Cross-account CloudTrail registro

Práticas recomendadas, considerações e limitações do Lake Formation

Use esta seção para encontrar rapidamente as práticas recomendadas, considerações e limitações no AWS Lake Formation.

Consulte Service Quotas para conhecer o número máximo de recursos de serviço ou operações da Conta da AWS.

Tópicos

- Práticas recomendadas e considerações sobre compartilhamento de dados entre contas
- Limitações de perfis vinculadas a serviços
- Limitações de acesso aos dados entre regiões
- Considerações e limitações das visualizações do catálogo de dados
- Limitações de filtragem de dados
- Considerações e limitações do modo de acesso híbrido
- Limitações para trazer dados do armazém de dados do Amazon Redshift para o AWS Glue Data Catalog
- Limitações da integração do catálogo do S3 Tables
- Considerações e limitações do compartilhamento de dados de armazenamento de metadados do Hive
- Limitações do compartilhamento de dados do Amazon Redshift
- Limitações da integração com o Centro de Identidade do IAM
- Considerações e práticas recomendadas de controle de acesso com base em tags do Lake Formation
- Considerações sobre controle de acesso por atributo, limitações e regiões aceitas

Práticas recomendadas e considerações sobre compartilhamento de dados entre contas

Os recursos de várias contas do Lake Formation permitem que os usuários compartilhem com segurança lagos de dados distribuídos em várias AWS organizações ou diretamente com os

diretores do IAM em outra conta Contas da AWS, fornecendo acesso refinado aos metadados do Catálogo de Dados e aos dados subjacentes.

Pense nas seguintes práticas recomendadas ao usar o compartilhamento de dados entre contas do Lake Formation:

- Não há limite para o número de concessões de permissão do Lake Formation que você pode conceder aos diretores em sua própria AWS conta. No entanto, o Lake Formation usa a capacidade AWS Resource Access Manager (AWS RAM) para concessões entre contas que sua conta pode fazer com o método de recurso nomeado. Para maximizar a AWS RAM capacidade, siga estas práticas recomendadas para o método de recurso nomeado:
  - Use o novo modo de concessão entre contas (versão 3 e superior em Configurações de versão entre contas) para compartilhar um recurso com um externo Conta da AWS. Para obter mais informações, consulte Como atualizar as configurações da versão de compartilhamento de dados entre contas.
  - Organize AWS contas em organizações e conceda permissões a organizações ou unidades organizacionais. Uma concessão para uma organização ou unidade organizacional conta como apenas uma concessão.

    A concessão para organizações ou unidades organizacionais também elimina a necessidade de aceitar um convite AWS Resource Access Manager (AWS RAM) de compartilhamento de recursos para a concessão. Para obter mais informações, consulte Acessar e visualizar tabelas e bancos de dados compartilhados do catálogo de dados.
  - Em vez de conceder permissões em várias tabelas individuais do banco de dados, use o curinga especial Todas as tabelas para conceder permissões em todas as tabelas do banco de dados. A concessão em Todas as tabelas conta como uma única concessão. Para obter mais informações, consulte Conceder permissões nos recursos do Catálogo de Dados.

  > **Note**
  >
  > Para obter mais informações sobre como solicitar um limite maior para o número de compartilhamentos de recursos em AWS RAM, consulte cotas AWS de serviço no. Referência geral da AWS

- Você deve criar um link de recurso para um banco de dados compartilhado para que esse banco de dados apareça nos editores de consulta Amazon Athena e no Amazon Redshift Spectrum. Da mesma forma, para poder consultar tabelas compartilhadas usando o Athena e o Redshift

Spectrum, você deve criar links de recursos para as tabelas. Em seguida, os links de recursos aparecem na lista de tabelas dos editores de consulta.

Em vez de criar links de recursos para várias tabelas individuais para consulta, você pode usar o curinga Todas as tabelas para conceder permissões em todas as tabelas em um banco de dados. Em seguida, ao criar um link de recurso para esse banco de dados e selecionar esse link de recurso de banco de dados no editor de consultas, você terá acesso a todas as tabelas desse banco de dados para sua consulta. Para obter mais informações, consulte Criação de links de recursos.

- Quando você compartilha recursos diretamente com principais em outra conta, a entidade principal do IAM na conta do destinatário pode não ter permissão para criar links de recursos para poder consultar as tabelas compartilhadas usando o Athena e o Amazon Redshift Spectrum. Em vez de criar um link de recurso para cada tabela compartilhada, o administrador do data lake pode criar um banco de dados provisório e conceder a permissão `CREATE_TABLE` ao grupo `ALLIAMPrincipal`. Em seguida, todas as entidades principais do IAM na conta do destinatário podem criar links de recursos no banco de dados de espaços reservados, e começar a consultar as tabelas compartilhadas.

  Veja o exemplo de comando CLI para conceder permissões para `ALLIAMPrincipals` em Conceder permissões de banco de dados usando o método de recurso nomeado.
- Quando as permissões entre contas são concedidas diretamente a uma entidade principal, somente o destinatário da concessão pode visualizar essas permissões. O administrador do data lake na conta da AWS do destinatário não pode visualizar essas concessões diretas.
- O Athena e o Redshift Spectrum oferecem suporte ao controle de acesso em nível de coluna, mas somente para inclusão, não exclusão. O controle de acesso em nível de coluna não é suportado em trabalhos de ETL no AWS Glue.
- Quando um recurso é compartilhado com sua AWS conta, você pode conceder permissões sobre o recurso somente aos usuários da sua conta. Você não pode conceder permissões sobre o recurso para outras AWS contas, para organizações (nem mesmo para sua própria organização) ou para o `IAMAllowedPrincipals` grupo.
- Não é possível conceder `DROP` ou `Super` em um banco de dados a uma conta externa.
- Revogue as permissões entre contas antes de excluir um banco de dados ou uma tabela. Caso contrário, você deverá excluir compartilhamentos de recursos órfãos em. AWS Resource Access Manager

Consulte também

- Considerações e práticas recomendadas de controle de acesso com base em tags do Lake Formation
- `CRIAR_TABELA` na seção Referência de permissões do Lake Formation para obter mais regras e limitações de acesso entre contas.

Limitações de perfis vinculadas a serviços

Uma função vinculada ao serviço é um tipo especial de função do IAM vinculada diretamente a. AWS Lake Formation Essa função tem permissões predefinidas que permitem que a Lake Formation realize ações em seu nome em todos AWS os serviços.

As limitações a seguir se aplicam ao usar um perfil vinculado a serviços (SLR) para registrar locais de dados no Lake Formation.

- Não é possível modificar políticas de perfil vinculado a serviços depois de criadas.
- Um perfil vinculado a serviços não aceita o compartilhamento criptografado de recursos de catálogo entre contas. Os recursos criptografados exigem permissões de AWS KMS chave específicas. Os perfis vinculados a serviços têm permissões predefinidas que não incluem a capacidade de trabalhar com recursos de catálogo criptografados em todas as contas.
- Ao registrar vários locais do Amazon S3, o uso do perfil vinculado a serviços pode fazer com que você exceda rapidamente os limites da política do IAM. Isso acontece porque, com funções vinculadas a serviços, AWS escreve a política para você e ela é incrementada como um grande bloco que inclui todos os seus registros. Você pode criar políticas gerenciadas pelo cliente com maior eficiência, distribuir permissões em várias políticas ou usar perfis diferentes para regiões distintas.
- O Amazon EMR no EC2 não consegue acessar dados a menos que você registre os locais de dados com perfis vinculados a serviços.
- As operações de função vinculadas ao serviço ignoram suas AWS políticas de controle de serviços.
- Quando você registra locais de dados com um perfil vinculado a serviços, ele atualiza as políticas do IAM com consistência eventual. Para acessar mais informações, consulte a documentação de solução de problemas do IAM no Guia do usuário do IAM.

- Não é possível definir `SET_CONTEXT = TRUE` nas configurações de data lake do Lake Formation ao usar perfis vinculados a serviços e está usando o Centro de Identidade do IAM. O motivo é que os perfis vinculados a serviços têm políticas de confiança imutáveis que são incompatíveis com a propagação de identidade confiável necessária para a auditoria `SetContext` com as entidades principais do Centro de Identidade do IAM.

Limitações de acesso aos dados entre regiões

O Lake Formation aceita a consulta de tabelas do catálogo de dados entre Regiões da AWS. Você pode acessar dados em uma região de outras regiões usando o Amazon Athena Amazon EMR e o AWS Glue ETL criando links de recursos em outras regiões apontando para os bancos de dados e tabelas de origem. Com o acesso à tabela entre regiões, você pode acessar dados entre regiões sem copiar os dados subjacentes ou os metadados no catálogo de dados.

As limitações a seguir se aplicam ao acesso a tabelas entre regiões.

- O Lake Formation não suporta a consulta de tabelas do catálogo de dados de outra região usando o Amazon Redshift Spectrum.
- No console do Lake Formation, as visualizações do banco de dados e da tabela não mostram os nomes dos bancos de dados/tabelas da região de origem.
- Para exibir a lista de tabelas em um banco de dados compartilhado de outra região, você precisa primeiro criar um link de recurso para o banco de dados compartilhado, depois selecionar o link do recurso e escolher Exibir tabelas.
- O Lake Formation não é compatível com chamadas de links de recursos entre regiões.
- O recurso entre regiões do Lake Formation não envolve cobranças adicionais pelas transferências de dados.

Considerações e limitações das visualizações do catálogo de dados

Estas são considerações e limitações que se aplicam às visualizações do catálogo de dados.

- Não é possível criar uma visualização do Data Catalog no console do Lake Formation. Você pode criar visualizações usando o AWS CLI ou SDK.

- É possível criar visualizações do Data Catalog em dez tabelas. Trata-se de um limite fixo. As tabelas de referência subjacentes de uma visualização podem pertencer ao mesmo banco de dados ou a bancos de dados diferentes na mesma AWS conta.
- Para outras considerações e limitações específicas sobre a criação de visualizações do Data Catalog usando o Redshift, consulte a seção Considerações e limitações das visualizações do Data Catalog no Guia do desenvolvedor do banco de dados Amazon Redshift. Para o Athena, consulte a seção Considerações e limitações das visualizações do Data Catalog no Guia do usuário do Amazon Athena.
- Você pode criar visualizações do Catálogo de Dados em tabelas registradas no Lake Formation no modo de acesso híbrido e no modo do Lake Formation.

  Ao usar as visualizações do Catálogo de Dados com o modo de acesso híbrido do Lake Formation, é recomendável garantir que as entidades principais de consumo da visualização aceitem as permissões do Lake Formation para as tabelas de base referidas na visualização sem conceder acesso. Isso garante que as tabelas base não sejam reveladas aos consumidores por meio de permissões AWS Glue do IAM.
- Não há restrições na versão de compartilhamento entre contas para compartilhar visualizações.
- As visualizações são versionadas da mesma forma que as tabelas do Catálogo de Dados quando você usa a instrução `ALTER VIEW` para um dialeto de visualização já criado. Não é possível reverter para uma visualização anterior porque a versão da visualização muda com as alterações dos dados subjacentes. Você pode excluir uma versão de visualização e ela será padronizada para a próxima versão mais recente disponível. Ao alterar a versão da visualização, seus dados devem estar sincronizados com o esquema da versão de visualização selecionada.
- Nenhum novo catálogo APIs de dados foi introduzido. Os existentes`CreateTable`,`UpdateTable`, `DeleteTable` e `GetTable` APIs são atualizados.
- O Amazon Redshift sempre cria visualizações com colunas varchar com base em tabelas com strings. É necessário converter colunas de string em varchar com um tamanho explícito ao adicionar dialetos de outros mecanismos.
- Conceder permissões de data lake a `All tables` em um banco de dados resultará em permissões do favorecido em todas as tabelas e visualizações do banco de dados.
- Não é possível criar visualizações:
  - Isso faz referência a outras visualizações.
  - Quando a tabela de referência é um link de recurso.
  - Quando a tabela de referência está em outra conta.

  - De metastores externos do Hive.
- As funções de definição de contas cruzadas não são suportadas nas visualizações do Redshift Spectrum Dialect.
- Não há suporte para links de recursos para o dialeto Athena no editor de consultas do Athena. Para usar funções de definição de contas cruzadas para o dialeto Athena, adicione a conta que hospeda as tabelas base como uma fonte de dados no Athena.

Limitações de filtragem de dados

Ao conceder permissões do Lake Formation em uma tabela do catálogo de dados, você pode incluir especificações de filtragem de dados para restringir o acesso a determinados dados nos resultados da consulta e nos mecanismos integrados ao Lake Formation. O Lake Formation usa a filtragem de dados para obter segurança por coluna, segurança por linha e segurança por célula. Será possível definir e aplicar filtros de dados em colunas aninhadas se os dados de origem contiverem estruturas aninhadas.

Notas e restrições para filtragem em nível de coluna

Há três maneiras de especificar a filtragem de colunas:

- Usando filtros de dados
- Usando filtragem de colunas simples ou filtragem de colunas aninhada.
- Usando TAGs.

A filtragem simples de colunas apenas especifica uma lista de colunas a serem incluídas ou excluídas. Tanto o console do Lake Formation quanto a API AWS CLI oferecem suporte à filtragem simples de colunas. Para ver um exemplo, consulte Grant with Simple Column Filtering.

As seguintes notas e restrições se aplicam à filtragem de colunas:

- AWS Glue 5.0 ou superior suporta controle de acesso refinado via Lake Formation somente para tabelas Apache Hive e Apache Iceberg.
- Para conceder com a opção `SELECT` e a filtragem de colunas, você deve usar uma lista de inclusão, não uma lista de exclusão. Sem a opção de concessão, você pode usar listas de inclusão ou exclusão.

- Para conceder `SELECT` em uma tabela com filtragem de colunas, você deve ter recebido a opção de concessão `SELECT` na tabela e sem nenhuma restrição de linha. Você deve ter acesso a todas as linhas.
- Se você conceder com a opção `SELECT` e a filtragem de colunas a uma entidade principal em sua conta, essa entidade principal deverá especificar a filtragem de colunas para as mesmas colunas ou um subconjunto das colunas concedidas ao conceder a outra entidade principal. Se você conceder com a opção `SELECT` e a filtragem de colunas a uma conta externa, o administrador do data lake na conta externa poderá conceder `SELECT` a todas as colunas a outra entidade principal em sua conta. No entanto, mesmo com todas as colunas com `SELECT`, esse entidade principal terá visibilidade somente nas colunas concedidas à conta externa.
- Você não pode aplicar a filtragem de colunas nas chaves de partição.
- Uma entidade principal com a permissão `SELECT` em um subconjunto de colunas em uma tabela não pode receber a permissão `ALTER`, `DROP`, `DELETE`, ou `INSERT` nessa tabela. Para uma entidade principal com a permissão `ALTER`, `DROP`, `DELETE`, ou `INSERT` em uma tabela, se você conceder a permissão `SELECT` com a filtragem de colunas, ela não terá efeito.

As seguintes notas e restrições se aplicam à filtragem de colunas aninhadas:

- É possível incluir ou excluir cinco níveis de campos aninhados em um filtro de dados.

  Example

  Col1.Col1_1.Col1_1_1.Col1_1_1_1.Col1_1_1_1_1
- Não é possível aplicar a filtragem de colunas em campos aninhados em colunas de partição.
- Se o esquema da tabela contiver um nome de coluna de nível superior (“cliente”.” address”) que tem o mesmo padrão de representação de campo aninhado em um filtro de dados (uma coluna aninhada com um nome de coluna de nível superior `customer` e um nome de campo aninhado `address` é especificado como `"customer"."address"` em um filtro de dados), você não pode especificar explicitamente o acesso à coluna de nível superior ou ao campo aninhado porque ambos são representados usando o mesmo padrão nas listas. inclusion/exclusion Isso é ambíguo, e o Lake Formation não poderá resolver se você estiver especificando a coluna de nível superior ou o campo aninhado.
- Se uma coluna de nível superior ou um campo aninhado contiver aspas duplas no nome, será necessário incluir uma segunda aspa dupla ao especificar o acesso a um campo aninhado na lista de inclusão e exclusão de um filtro de células de dados.

Example

Exemplo de nome de coluna aninhada com aspas duplas: `a.b.double"quote`

Example

Exemplo de representação de coluna aninhada em um filtro de dados: `"a"."b"."double""quote"`

Limitações de filtragem no nível de célula

Tenha em mente as seguintes notas e restrições para filtragem em nível de linha e de célula:

- Em colunas aninhadas, visualizações e links de recurso, não é possível usar a segurança no nível de célula.
- Todas as expressões aceitas em colunas de nível superior também são aceitas em colunas aninhadas. No entanto, os campos aninhados em colunas de partição NÃO devem ser referenciados ao definir expressões aninhadas em nível de linha.
- A segurança por célula está disponível em todas as regiões ao usar o Athena Engine versão 3 ou o Amazon Redshift Spectrum. Para outros serviços, a segurança por célula só está disponível nas regiões mencionadas em Regiões aceitas.
- As instruções `SELECT INTO` não são compatíveis.
- Os tipos de dados `array` e `map` não são compatíveis com expressões de filtro de linha. O tipo de dados `struct` é aceito.
- Não há limite para o número de filtros de dados que podem ser definidos em uma tabela, mas há um limite de cem filtros de dados para uma única entidade principal em uma tabela.
- Para aplicar um filtro de dados com uma expressão de filtro de linha, você deve ter a opção `SELECT` de concessão em todas as colunas da tabela. Essa restrição não se aplica a administradores em contas externas quando a concessão foi feita para a conta externa.
- Se uma entidade principal for membro de um grupo e tanto a entidade principal quanto o grupo receberem permissões em um subconjunto de linhas, as permissões de linha efetivas da entidade principal são a união das permissões do entidade principal e das permissões do grupo.
- Os seguintes nomes de colunas são restritos em uma tabela para filtragem em nível de linha e de célula:
  - ctid

- oid
- xmin
- cmin
- xmax
- cmax
- tableoid
- insertxid
- deletexid
- importoid
- redcatuniqueid

- Se você aplicar a expressão de filtro de todas as linhas em uma tabela simultaneamente com outras expressões de filtro com predicados, a expressão de todas as linhas prevalecerá sobre todas as outras expressões de filtro.
- Quando as permissões em um subconjunto de linhas são concedidas a uma AWS conta externa e o administrador do data lake da conta externa concede essas permissões a um principal nessa conta, o predicado de filtro efetivo do principal é a interseção do predicado da conta com qualquer predicado que tenha sido concedido diretamente ao principal.

  Por exemplo, se a conta tiver permissões de linha com o predicado `dept='hr'` e a entidade principal tiver recebido a permissão `country='us'` separadamente, a entidade principal terá acesso somente às linhas com `dept='hr'` e `country='us'`.

Para obter mais informações sobre a filtragem em nível de célula, consulte Filtragem de dados e segurança por célula no Lake Formation.

Para analisar considerações e limitações ao consultar tabelas usando o Amazon Redshift Spectrum com políticas de segurança em nível de linha, consulte Considerações e limitações ao usar políticas de RLS no Guia do desenvolvedor do banco de dados Amazon Redshift.

Considerações e limitações do modo de acesso híbrido

O modo de acesso híbrido oferece a flexibilidade de habilitar seletivamente as permissões do Lake Formation para bancos de dados e tabelas no seu AWS Glue Data Catalog.

Com o modo de acesso híbrido, agora você tem um caminho incremental que permite definir permissões do Lake Formation para um conjunto específico de usuários sem interromper as políticas de permissão de outros usuários ou workloads existentes.

As considerações e limitações a seguir se aplicam ao modo de acesso híbrido.

Limitações

- Atualizar o registro de localização do Amazon S3 – Você não pode editar parâmetros de um local registrado no Lake Formation usando uma função vinculada ao serviço.
- Opção de ativação ao usar tags do LF – Quando você pode conceder permissões do Lake Formation usando tags do LF, você pode optar por entidades principais para aplicar as permissões do Lake Formation em uma etapa consecutiva, escolhendo bancos de dados e tabelas com tags do LF anexadas.
- Acesso ao modo de acesso híbrido: o acesso ao modo de acesso híbrido no Lake Formation é limitado a usuários com permissões de administrador do data lake ou administrador somente leitura.
- Optar por entidades principais – Atualmente, somente uma função de administrador de data lake pode optar por entidades principais para recursos.
- Aceitar todas as tabelas em um banco de dados: em concessões entre contas, ao conceder permissões e aceitar todas as tabelas em um banco de dados, é necessário aceitar o banco de dados também para que as permissões funcionem.

Considerações

- Atualização da localização do Amazon S3 registrada no Lake Formation para o modo de acesso híbrido – Não recomendamos converter uma localização de dados do Amazon S3 que já esteja registrada no Lake Formation para o modo de acesso híbrido, embora isso possa ser feito.
- Comportamentos da API quando um local de dados é registrado no modo de acesso híbrido
  - CreateTable — O local é considerado registrado no Lake Formation, independentemente da bandeira do modo de acesso híbrido e do status de opt-in. Assim, o usuário precisa da permissão de localização dos dados para criar uma tabela.
  - CreatePartition/BatchCreatePartition/UpdatePartition(quando a localização da partição é atualizada para apontar para a localização registrada com híbrido) — A localização do Amazon S3 é considerada registrada no Lake Formation, independentemente da bandeira do modo de acesso híbrido e do status de aceitação.

- CreateDatabase/UpdateDatabase (quando a localização do banco de dados é atualizada para apontar para a localização registrada no modo de acesso híbrido) — A localização é considerada registrada no Lake Formation, independentemente da bandeira do modo de acesso híbrido e do status de ativação. Assim, o usuário precisa da permissão de localização de dados para criar ou atualizar um banco de dados.
- UpdateTable (quando a localização de uma tabela é atualizada para apontar para a localização registrada no modo de acesso híbrido) — A localização é considerada registrada no Lake Formation, independentemente da bandeira do modo de acesso híbrido e do status de ativação. Assim, o usuário precisa de permissão de localização de dados para atualizar a tabela. Se a localização da tabela não for atualizada ou estiver apontando para uma localização que não esteja registrada no Lake Formation, o usuário não precisará da permissão de localização de dados para atualizar a tabela.

Limitações para trazer dados do armazém de dados do Amazon Redshift para o AWS Glue Data Catalog

É possível catalogar e gerenciar o acesso aos dados de analytics nos data warehouses do Amazon Redshift usando o AWS Glue Data Catalog. As limitações a seguir se aplicam a:

- O compartilhamento entre contas não é aceito em nível de catálogo federado. Contudo, você pode compartilhar bancos de dados e tabelas individuais de dentro de um catálogo federado entre Conta da AWS s.
- Você deve ter as configurações de versão da conta cruzada versão 4 ou superior para compartilhar bancos de dados ou tabelas no catálogo federado entre Conta da AWS s.
- O Data Catalog comporta a criação somente de catálogos de nível superior.
- Você só pode atualizar a descrição de catálogos no Redshift Managed Storage (RMS).
- A configuração de permissões em catálogos federados, bem como em bancos de dados e tabelas no catálogo federado para o grupo `IAMAllowedPrincipals` não é aceita.

- As operações de linguagem de definição de dados (DDL) no catálogo por meio de mecanismos, como Athena, Amazon EMR Spark ou outros, incluindo a definição de configurações do catálogo, não são aceitas.
- Operações de DDL em tabelas do RMS usando o Athena não são aceitas.

- A criação de visualizações materializadas não é suportada, seja por meio do Athena, do Apache Spark, AWS Glue Data Catalog do ou do consumidor do Amazon Redshift.
- O Athena não aceita experiência de vários catálogos. Ele só consegue se conectar a um único catálogo específico por vez. O Athena não consegue acessar nem consultar vários catálogos simultaneamente.
- As operações de marcação e ramificação em tabelas do Iceberg por meio do Athena e do Amazon Redshift não são aceitas.
- Time Travel em tabelas do RMS não é aceito.
- Catálogos de vários níveis com tabelas de data lake não são aceitos. Todos os dados armazenados no Amazon S3 para uso com tabelas de data lake devem residir no padrão AWS Glue Data Catalog e não podem ser organizados em catálogos de vários níveis.
- No Amazon Redshift, as unidades de compartilhamento de dados não são adicionadas ao namespace registrado. Clusters e namespaces são sinônimos. Depois de publicar um cluster no AWS Glue Data Catalog, você não pode adicionar novos dados.
- O Amazon EMR no EC2 não comporta a união entre tabelas do RMS e Tabelas do Amazon S3. Somente o EMR Sem Servidor aceita esse recurso.
- Esquemas e tabelas externos não são aceitos.
- As tabelas do RMS são acessíveis somente por meio do endpoint de extensão no Catálogo REST do AWS Glue Iceberg.
- As tabelas do Hive não podem ser acessadas por mecanismos de terceiros conectados ao Catálogo REST do AWS Glue Iceberg.
- O nível de isolamento read_committed em tabelas do RMS por meio do Spark será aceito.
- Os nomes dos bancos de dados do Redshift são tratados como sem distinção entre maiúsculas e minúsculas no AWS Glue Data Catalog, restritos a 128 caracteres e podem ser alfanuméricos com traços (-) e sublinhados (_).
- Os nomes dos catálogos não fazem distinção de maiúsculas e minúsculas, são restritos a 50 caracteres e podem ser alfanuméricos com traços (-) e sublinhados (_).
- O Amazon Redshift não aceita o uso dos comandos GRANT e REVOKE no estilo SQL do Lake Formation para gerenciar permissões de acesso em tabelas publicadas no AWS Glue Data Catalog.
- As políticas de segurança por linha e de mascaramento dinâmico de dados anexadas ao cluster produtor (de origem) do Amazon Redshift não serão aplicadas. Em vez disso, as permissões de acesso definidas no Lake Formation serão aplicadas aos dados compartilhados.

- Operações de linguagem de definição de dados (DDL) e linguagem de manipulação de dados (DML) em links de tabela não são aceitas.
- Se as palavras-chave reservadas não forem devidamente tratadas com caracteres de escape, ocorrerão falhas ou erros.
- A criptografia de dados em cenários de vários catálogos não é aceita.

Limitações da integração do catálogo do S3 Tables

O Amazon S3 Tables se integra ao AWS Glue Data Catalog (Catálogo de dados) e registra o catálogo como um local de dados do Lake Formation. Você pode configurar esse registro no console do Lake Formation ou usando o serviço APIs.

> **Note**
>
> Se você tiver uma política baseada em recursos do IAM ou do S3 Tables que restringe os usuários do IAM e as funções do IAM com base nas tags principais, você deve anexar as mesmas tags principais à função do IAM que Lake Formation usa para acessar seus dados do S3 (por exemplo, LakeFormationDataAccessRole) e conceder a essa função as permissões necessárias. Isso é necessário para que a política de controle de acesso baseada em tags funcione corretamente com integração de analytics da funcionalidade Tabelas do S3.

As seguintes limitações se aplicam à integração do catálogo de tabelas do S3 com o Data Catalog e o Lake Formation:

- AWS Glue e Lake Formation não oferecem suporte a nomes de colunas com letras maiúsculas e minúsculas e convertem todos os nomes de colunas em minúsculas. É necessário verificar se os nomes das colunas da tabela são exclusivos quando convertidos em minúsculas. Use `customer_id` em vez de `customerId`. O uso de nomes de colunas com letras maiúsculas e minúsculas foi aceito somente durante a versão prévia.
- A CreateCatalog API não pode criar buckets de tabela no Amazon S3.
- A SearchTables API não pode pesquisar tabelas do S3.

Considerações e limitações do compartilhamento de dados de armazenamento de metadados do Hive

Com a federação de AWS Glue Data Catalog metadados (federação do catálogo de dados), você pode conectar o catálogo de dados a metastores externos que armazenam metadados para seus dados do Amazon S3 e gerenciar com segurança as permissões de acesso aos dados usando. AWS Lake Formation

As seguintes considerações e limitações se aplicam aos bancos de dados federados criados a partir dos bancos de dados do Hive:

Considerações

- AWS SAM suporte de aplicativos — Você é responsável pela disponibilidade dos recursos do aplicativo que são AWS SAM implantados (Amazon API Gateway e pela função Lambda). Certifique-se de que a conexão entre o metastore AWS Glue Data Catalog e o Hive esteja funcionando quando os usuários executam consultas.
- Requisito da versão do metastore do Hive: é possível criar bancos de dados federados somente usando o Apache Hive versão 3 e posterior.
- Requisito de banco de dados mapeado — Todo banco de dados do Hive deve ser mapeado para um novo banco de dados no Lake Formation.
- Suporte à federação em nível de banco de dados — Você pode se conectar ao repositório do Hive somente no nível do banco de dados.
- Permissões em bancos de dados federados — As permissões aplicadas em um banco de dados federado ou tabelas em um banco de dados federado persistem mesmo quando uma tabela de origem ou um banco de dados é excluído. Quando o banco de dados ou tabela de origem são recriados, você não precisa conceder as permissões novamente. Quando uma tabela federada com permissões do Lake Formation é excluída na fonte, as permissões do Lake Formation ainda estão visíveis e você pode revogá-las se necessário.

  Se um usuário excluir um banco de dados federado, todas as permissões correspondentes serão perdidas. Recriar o mesmo banco de dados com o mesmo nome não recuperará as permissões do Lake Formation. Os usuários precisarão configurar novas permissões novamente.
- IAMAllowedPrincipais permissões de grupo em bancos de dados federados — Com base no`DataLakeSettings`, Lake Formation pode definir permissões para todos os bancos de dados e tabelas para um grupo virtual chamado`IAMAllowedPrincipal`. O `IAMAllowedPrincipal`

se refere a todos os diretores do IAM que têm acesso aos recursos do catálogo de dados por meio das políticas principais e políticas de AWS Glue recursos do IAM. Se essas permissões existirem em um banco de dados ou tabela, todos as entidades principais terão acesso ao banco de dados ou à tabela.

No entanto, o Lake Formation não aceita permissões `IAMAllowedPrincipal` em tabelas em bancos de dados federados. Ao criar bancos de dados federados, certifique-se de passar o parâmetro `CreateTableDefaultPermissions` como uma lista vazia.

Para obter mais informações, consulte Alterando as configurações padrão do seu data lake.

- Unir tabelas em consultas — Você pode unir tabelas de repositório do Hive com tabelas nativas do catálogo de dados para executar consultas.

Limitações

- Limitação na sincronização de metadados entre o AWS Glue Data Catalog e o metastore do Hive - Depois de estabelecer a conexão do metastore do Hive, você precisa criar um banco de dados federado para sincronizar os metadados no metastore do Hive com o. AWS Glue Data Catalog As tabelas no banco de dados federado são sincronizadas em runtime quando os usuários executam consultas.
- Limitação na criação de novas tabelas em um banco de dados federado — Você não poderá criar novas tabelas em bancos de dados federados.
- Limitação de permissão de dados — O suporte para permissões nas visualizações de tabela do Repositório do Hive não está disponível.

Limitações do compartilhamento de dados do Amazon Redshift

AWS Lake Formation permite que você gerencie dados com segurança em um compartilhamento de dados do Amazon Redshift. O Amazon Redshift é um serviço de armazém de dados totalmente gerenciado em escala de petabytes na nuvem. AWS Ao usar o recurso de compartilhamento de dados, o Amazon Redshift ajuda você a compartilhar dados entre Contas da AWS. Para obter mais informações sobre o compartilhamento de dados do Amazon Redshift, consulte Visão geral do compartilhamento de dados no Amazon Redshift.

As seguintes observações e restrições aplicam-se a bancos de dados federados criados a partir de unidades de compartilhamento de dados do Amazon Redshift:

- Requisito de banco de dados mapeado — Toda unidade de compartilhamento de dados do Amazon Redshift deve ser mapeada em um novo banco de dados no Lake Formation. Isso é necessário para manter nomes de tabela exclusivos quando a representação dos objetos da unidade de compartilhamento de dados é nivelada no banco de dados do catálogo de dados.
- Limitação na criação de novas tabelas em um banco de dados federado — Você não poderá criar novas tabelas em bancos de dados federados.
- Permissões nos bancos de dados federados — As permissões aplicadas em um banco de dados federado ou tabelas em um banco de dados federado persistem mesmo quando uma tabela de origem ou um banco de dados é excluído. Quando o banco de dados ou a tabela de origem são recriados, você não precisa conceder as permissões novamente. Quando uma tabela federada com permissões do Lake Formation é excluída na fonte, as permissões do Lake Formation ainda estarão visíveis e você poderá revogá-las se necessário.

  Se um usuário excluir um banco de dados federado, todas as permissões correspondentes serão perdidas. Recriar o mesmo banco de dados com o mesmo nome não recuperará as permissões do Lake Formation. Os usuários precisarão configurar novas permissões novamente.
- IAMAllowedPrincipais permissões de grupo em bancos de dados federados — Com base no`DataLakeSettings`, Lake Formation pode definir permissões para todos os bancos de dados e tabelas para um grupo virtual chamado`IAMAllowedPrincipal`. O `IAMAllowedPrincipal` se refere a todos os diretores do IAM que têm acesso aos recursos do catálogo de dados por meio das políticas principais e políticas de AWS Glue recursos do IAM. Se essas permissões existirem em um banco de dados ou tabela, todos as entidades principais terão acesso ao banco de dados ou à tabela.

  No entanto, o Lake Formation não aceita permissões `IAMAllowedPrincipal` em tabelas em bancos de dados federados. Ao criar bancos de dados federados, certifique-se de passar o parâmetro `CreateTableDefaultPermissions` como uma lista vazia.

  Para obter mais informações, consulte Alterando as configurações padrão do seu data lake.
- Filtragem de dados — No Lake Formation, você pode conceder permissões em uma tabela em um banco de dados federado com filtragem em nível de coluna e em nível de linha. No entanto, você não pode combinar a filtragem em nível de coluna e em nível de linha para restringir o acesso na granularidade em nível de célula em tabelas em bancos de dados federados.
- Identificador de distinção entre maiúsculas e minúsculas — Os objetos de unidades de compartilhamento de dados do Amazon Redshift gerenciados pelo Lake Formation suportarão nomes de tabelas e nomes de colunas somente em minúsculas. Não ative o identificador

de diferenciação de maiúsculas e minúsculas para bancos de dados, tabelas e colunas nas unidades de compartilhamento de dados do Amazon Redshift, caso eles sejam compartilhados e gerenciados usando o Lake Formation.

- Suporte a consultas: você pode consultar unidades de compartilhamento de dados do Amazon Redshift gerenciadas pelo Lake Formation com o Amazon Redshift. O Athena não aceita consultas a unidades de compartilhamento de dados do Amazon Redshift gerenciadas pelo Lake Formation.

Para obter mais informações sobre limitações ao trabalhar com unidades de compartilhamento de dados no Amazon Redshift, consulte Limitações do compartilhamento de dados no Guia do desenvolvedor de banco de dados do Amazon Redshift.

Limitações da integração com o Centro de Identidade do IAM

Com Centro de Identidade do AWS IAM, você pode se conectar a provedores de identidade (IdPs) e gerenciar centralmente o acesso de usuários e grupos em todos os serviços de AWS análise. Você pode configurar AWS Lake Formation como um aplicativo habilitado no IAM Identity Center, e os administradores do data lake podem conceder permissões refinadas a usuários e grupos autorizados sobre recursos. AWS Glue Data Catalog

As seguintes limitações se aplicam à integração do Lake Formation com o Centro de Identidade do IAM:

- Não é possível atribuir usuários e grupos do Centro de Identidade do IAM como administradores de data lake ou administradores somente leitura no Lake Formation.

  Usuários e grupos do IAM Identity Center podem consultar recursos criptografados do catálogo de dados se você estiver usando uma função do IAM que AWS Glue possa assumir em seu nome para criptografar e descriptografar o catálogo de dados. AWS as chaves gerenciadas não oferecem suporte à propagação de identidade confiável.
- Usuários e grupos do Centro de Identidade do IAM só podem invocar operações de API listadas na política `AWSIAMIdentityCenterAllowListForIdentityContext` fornecida pelo Centro de Identidade do IAM.
- O Lake Formation permite que os perfis do IAM de contas externas atuem como perfis operadores em nome dos usuários e grupos do Centro de Identidade do IAM para acessar os recursos do Catálogo de Dados, mas as permissões só podem ser concedidas em recursos do Catálogo de Dados dentro da conta proprietária. Se você tentar conceder permissões a usuários e grupos do

Centro de Identidade do IAM em recursos do Catálogo de Dados em uma conta externa, o Lake Formation vai gerar o seguinte erro: “Cross-account grants are not supported for the principal”.

- Ao usar o Lake Formation com o Centro de Identidade do IAM, a configuração de atribuição da aplicação é definida como `false` padrão. Se você modificar essa configuração diretamente por meio da API do Centro de Identidade do IAM, deverá gerenciar todas as atribuições da aplicação manualmente usando a API. O Lake Formation não sincroniza nem gerencia automaticamente as alterações de atribuição feitas fora de seus fluxos de trabalho padrão, o que pode afetar os padrões de acesso e os fluxos de autorização em seu ambiente de data lake.

Considerações e práticas recomendadas de controle de acesso com base em tags do Lake Formation

É possível criar, manter e atribuir tags do LF para controlar o acesso a bancos de dados, tabelas e colunas do catálogo de dados.

Pense nas seguintes práticas recomendadas ao usar o controle de acesso com base em tags do Lake Formation:

- Todas as tags do LF devem ser predefinidas antes de poderem ser atribuídas aos recursos do catálogo de dados ou concedidas às entidades principais.

  O administrador do data lake pode delegar tarefas de gerenciamento de tags gerando criadores de tags do LF com as permissões necessárias do IAM. Os engenheiros e analistas de dados decidem sobre as características e os relacionamentos das tags do LF. Os criadores da tags do LF então criam e mantêm as tags do LF no Lake Formation.
- Você pode atribuir várias tags do LF aos recursos do catálogo de dados. Somente um valor para uma chave específica pode ser atribuído a um recurso específico.

  Por exemplo, você pode atribuir `module=Orders`, `region=West` e `division=Consumer` e assim por diante a um banco de dados, uma tabela ou uma coluna. Você não pode atribuir `module=Orders,Customers`.
- Você não pode atribuir tags do LF aos recursos ao criá-los. Você só pode adicionar tags do LF aos recursos existentes.
- Você pode conceder expressões de tag do LF, não apenas tags do LF únicas, a uma entidade principal.

  Uma expressão de tag do LF se parece com a seguinte (em pseudocódigo).

```
module=sales AND division=(consumer OR commercial)
```

Uma entidade principal que recebe essa expressão de tag do LF pode acessar somente os recursos do catálogo de dados (bancos de dados, tabelas e colunas) que receberam `module=sales` e `division=consumer` ou `division=commercial`. Se você quiser que a entidade principal possa acessar recursos que tenham `module=sales` ou `division=commercial`, não inclua ambos na mesma concessão. Faça duas concessões, uma para `module=sales` e outra para `division=commercial`.

A expressão de tag do LF mais simples consiste em apenas uma tag do LF, como `module=sales`.

- Uma entidade principal que recebe permissões em uma tag do LF com vários valores pode acessar os recursos do catálogo de dados com qualquer um deles. Por exemplo, se um usuário receber uma tag do LF com chave = `module` e valores = `orders,customers`, o usuário terá acesso aos recursos atribuídos `module=orders` ou `module=customers`.
- Você precisa ter permissão `Grant with LF-Tag expressions` para conceder permissões de dados nos recursos do catálogo de dados usando o método LF-TBAC. O administrador do data lake e o criador da tag do LF recebem implicitamente essa permissão. Uma entidade principal que tenha a permissão `Grant with LFTag expressions` pode conceder permissões de dados sobre os recursos usando:
  - o método de recurso nomeado
  - o método LF-TBAC, mas usando apenas a mesma expressão de tag do LF

    Por exemplo, suponha que o administrador do data lake faça a seguinte concessão (em pseudocódigo).

    ```
    GRANT (SELECT ON TABLES) ON TAGS module=customers, region=west,south TO user1 WITH
     GRANT OPTION
    ```

    Nesse caso, `user1` pode conceder `SELECT` em tabelas a outras entidades principais usando o método LF-TBAC, mas somente com a expressão de tag do LF completa `module=customers, region=west,south`.
- Se uma entidade principal receber permissões em um recurso com o método LF-TBAC e o método de recurso nomeado, as permissões que a entidade principal tem sobre o recurso são a união das permissões concedidas pelos dois métodos.

- O Lake Formation oferece suporte à concessão de `DESCRIBE` e `ASSOCIATE` em tags do LF em todas as contas e à concessão de permissões nos recursos do catálogo de dados em todas as contas usando o método LF-TBAC. Em ambos os casos, o principal é o ID AWS da conta.

  > **Note**
  >
  > O Lake Formation aceita concessões entre contas para organizações e unidades organizacionais usando o método LF-TBAC. Para usar esse recurso, você precisa atualizar as Configurações de versão entre contas para a Versão 3.

  Para obter mais informações, consulte Cross-account compartilhamento de dados em Lake Formation.
- Os recursos do catálogo de dados criados em uma conta só podem ser marcados usando tags do LF criadas na mesma conta. As tags do LF criadas em uma conta não podem ser associadas a recursos compartilhados de outra conta.
- Usar o controle de acesso baseado em tags do Lake Formation (LF-TBAC) para conceder acesso entre contas aos recursos do Catálogo de Dados requer acréscimos à política de recursos do Catálogo de Dados para sua conta. AWS Para obter mais informações, consulte Pré-requisitos.
- As chaves da tag do LF e os valores da tag do LF não podem exceder 50 caracteres de comprimento.
- O número máximo de tags do LF que podem ser atribuídas a um recurso do catálogo de dados é 50.
- Os seguintes limites são limites flexíveis:
  - O número máximo de etiquetas LF que podem ser criadas é 10000.
  - O número máximo de valores que podem ser definidos para uma tag do LF é 1000.
- As chaves e valores das tags são convertidos em letras minúsculas quando são armazenados.
- Somente um valor para uma tag do LF pode ser atribuído a um recurso específico.
- Se várias tags do LF forem concedidas a uma entidade principal com uma única concessão, a entidade principal poderá acessar somente os recursos do catálogo de dados que tenham todas as tags do LF.
- Se uma avaliação da expressão de tag do LF resultar em acesso somente a um subconjunto de colunas da tabela, mas a permissão Lake Formation concedida quando há uma correspondência for uma das permissões que exigiram acesso total à coluna, ou seja, `Alter`, `Drop`, `Insert` ou `Delete`, nenhuma dessas permissões será concedida. Em vez disso, somente `Describe` é

concedido. Se a permissão concedida for `All` (`Super`), somente `Select` e `Describe` serão concedidas.

- Curingas não são usados com tags do LF. Para atribuir uma tag do LF a todas as colunas de uma tabela, atribua a tag do LF à tabela e todas as colunas na tabela herdam a tag do LF. Para atribuir uma tag do LF a todas as tabelas em um banco de dados, atribua a tag do LF ao banco de dados, e todas as tabelas no banco de dados herdam essa tag do LF.
- É possível criar até mil expressões de tag LF em uma conta.
- Você pode usar até cinquenta expressões de tag LF para conceder permissões a uma entidade principal nos recursos do Data Catalog.
- Quando você concede ou revoga permissões em uma expressão de tag LF em linha, o tamanho da expressão de tag LF não pode exceder 900 bytes. Para conceder permissões em expressões de tag LF maiores, use as expressões de tag LF salvas. Para obter mais informações, consulte Criação de LF-Tag expressões.
- Para adicionar o LF-Tag aos catálogos federados existentes do Redshift que foram criados antes do lançamento geral do suporte ao LF-Tag para catálogos federados, você precisa entrar em contato com a equipe de suporte para obter assistência. AWS

Considerações sobre controle de acesso por atributo, limitações e regiões aceitas

As considerações e limitações a seguir se aplicam ao controle de acesso por atributo (ABAC).

- O ABAC não comporta a concessão de acesso usando políticas de tag LF.
- As permissões que podem ser concedidas não estão disponíveis com o ABAC.
- O ABAC não comporta a concessão de permissões aos usuários do Centro de Identidade do IAM.
- Quando você usa concessões ABAC em uma tabela no Lake Formation, este não concede permissões `DESCRIBE` ao banco de dados ou catálogo principal. Isso difere dos cenários não ABAC, nos quais o Lake Formation concede permissões `DESCRIBE` implícitas aos recursos principais.
- Todas as entidades principais com a chave de tag `AmazonDataZoneProject` são sempre tratadas como aceitas pelo Lake Formation para todos os recursos do Data Catalog.
- O ABAC suporta somente atributos de string.

Solução de problemas do Lake Formation

Se você encontrar problemas ao trabalhar com o AWS Lake Formation, consulte os tópicos desta seção.

Tópicos

- Solução de problemas gerais
- Resolução de problemas de acesso entre contas
- Solução de problemas em esquemas e fluxos de trabalho
- Problemas conhecidos do AWS Lake Formation
- Mensagem de erro atualizada

Solução de problemas gerais

Use as informações aqui para ajudá-lo a diagnosticar e corrigir vários problemas do Lake Formation.

Erro: permissões insuficientes do Lake Formation em <Amazon S3 location>

Foi feita uma tentativa de criar ou alterar um recurso do catálogo de dados sem permissões de localização de dados na localização do Amazon S3 apontada pelo recurso.

Se um banco de dados ou tabela do Data Catalog apontar para uma localização do Amazon S3, ao conceder as permissões do Lake Formation `CREATE_TABLE` ou `ALTER`, você também deverá conceder a permissão `DATA_LOCATION_ACCESS` no local. Ao conceder permissões a contas ou organizações externas, é necessário incluir a opção de concessão.

Depois que essas permissões forem concedidas a uma conta externa, o administrador do data lake dessa conta deverá conceder as permissões às entidades principais (usuários ou funções) na conta. Ao conceder a `DATA_LOCATION_ACCESS` permissão recebida de outra conta, você deve especificar a ID do catálogo (ID da AWS conta) da conta do proprietário. A conta do proprietário é a conta que registrou o local.

Para obter mais informações, consulte Controle de acesso a dados subjacente e Conceder permissões de localização de dados.

Erro: “permissões de chave de criptografia insuficientes para a API Glue”

Foi feita uma tentativa de conceder permissões do Lake Formation sem permissões AWS Identity and Access Management (IAM) na chave de AWS KMS criptografia de um catálogo de dados criptografado.

Minha consulta Amazon Athena ou do Amazon Redshift que usa manifestos está falhando

O Lake Formation não suporta consultas que usam manifestos.

Erro: "permissão(ões) do Lake Formation insuficiente(s): necessária a criação de tag no catálogo"

Ele user/role deve ser um administrador de data lake.

Erro ao excluir administradores de data lake inválidos

É necessário excluir todos os administradores de data lake inválidos (perfis do IAM excluídos que são definidos como administradores de data lake) simultaneamente. Se você tentar excluir separadamente administradores de data lake inválidos, o Lake Formation gerará um erro de entidade principal inválida.

Erro: “O chamador está temporariamente não autorizado a realizar esta operação no recurso...”

Foi feita uma tentativa de consultar uma tabela usando um link de recurso que pode ter sido redirecionado. Para evitar esse erro, seu aplicativo deve adotar o QuerySessionContextprotocolo. Consulte o Lake Formation fluxo de trabalho para operações de API de integração de aplicativos para obter mais informações. Caso contrário, aguarde 24 horas antes de tentar executar sua consulta novamente.

Resolução de problemas de acesso entre contas

Use estas informações para ajudar a diagnosticar e corrigir problemas de acesso entre contas.

Tópicos

- Eu concedi uma permissão para várias contas do Lake Formation, mas o destinatário não consegue ver o recurso
- As entidades principais da conta do destinatário podem ver o recurso do catálogo de dados, mas não podem acessar os dados subjacentes
- Erro: “Falha na associação porque o chamador não foi autorizado” ao aceitar um convite de compartilhamento AWS RAM de recursos
- Erro: “não autorizado a conceder permissões para o recurso”
- Erro: “Acesso negado para recuperar informações AWS da organização”
- Erro: “organização <organization-ID> não encontrada”
- Erro: "permissões insuficientes do Lake Formation: combinação ilegal"
- ConcurrentModificationException em grant/revoke solicitações para contas externas
- Erro ao usar o Amazon EMR para acessar dados compartilhados por meio de várias contas

Eu concedi uma permissão para várias contas do Lake Formation, mas o destinatário não consegue ver o recurso

- O usuário na conta do destinatário é administrador do data lake? Somente administradores de data lake podem ver o recurso no momento do compartilhamento.
- Você está compartilhando com uma conta externa à sua organização usando o método de recurso nomeado? Nesse caso, o administrador do data lake da conta do destinatário deve aceitar um convite de compartilhamento de recursos em AWS Resource Access Manager (AWS RAM).

  Para obter mais informações, consulte the section called “Aceitando um convite AWS RAM de compartilhamento de recursos”.
- Você está usando políticas de recursos em nível de conta (catálogo de dados) no AWS Glue? Se sim, se você usar o método de recursos nomeados, deverá incluir uma declaração especial na política que autorize a AWS RAM o compartilhamento de políticas em seu nome.

  Para obter mais informações, consulte the section called “Gerenciando permissões entre contas usando ambos AWS Glue e Lake Formation”.
- Você tem as permissões AWS Identity and Access Management (IAM) necessárias para conceder acesso entre contas?

  Para obter mais informações, consulte the section called “Pré-requisitos”.

- O recurso para o qual você concedeu permissões não deve ter nenhuma permissão do Lake Formation concedida ao grupo `IAMAllowedPrincipals`.
- Há uma declaração `deny` sobre o recurso na política em nível de conta?

As entidades principais da conta do destinatário podem ver o recurso do catálogo de dados, mas não podem acessar os dados subjacentes

Os diretores da conta do destinatário devem ter as permissões necessárias AWS Identity and Access Management (IAM). Para obter detalhes, consulte Como acessar os dados subjacentes de uma tabela compartilhada.

Erro: “Falha na associação porque o chamador não foi autorizado” ao aceitar um convite de compartilhamento AWS RAM de recursos

Depois de conceder acesso a um recurso em uma conta diferente, quando a conta receptora tenta aceitar o convite de compartilhamento de recursos, a ação falha.

```
$ 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
        }
    ]
}
```

O erro ocorre porque `glue:PutResourcePolicy` é invocado pelo AWS Glue quando a conta receptora aceita o convite de compartilhamento de recursos. Para resolver o problema, permita a `glue:PutResourcePolicy` ação da função assumida usada pela producer/grantor conta.

Erro: “não autorizado a conceder permissões para o recurso”

Foi feita uma tentativa de conceder permissões entre contas em um banco de dados ou tabela pertencente a outra conta. Quando um banco de dados ou tabela é compartilhado com sua conta, como administrador do data lake, você pode conceder permissões sobre ele somente aos usuários da sua conta.

Erro: “Acesso negado para recuperar informações AWS da organização”

Sua conta é uma conta de gerenciamento da AWS Organizations e você não tem as permissões necessárias para recuperar informações da organização, como unidades organizacionais na conta.

Para obter mais informações, consulte Required permissions for cross-account grants.

Erro: “organização <organization-ID> não encontrada”

Foi feita uma tentativa de compartilhar um recurso com uma organização, mas o compartilhamento com organizações não está habilitado. Habilitar o compartilhamento de recursos com organizações.

Para obter mais informações, consulte Enable Sharing with AWS Organizations no Guia AWS RAM do Usuário.

Erro: "permissões insuficientes do Lake Formation: combinação ilegal"

Um usuário compartilhou um recurso do catálogo de dados enquanto as permissões do Lake Formation foram concedidas ao grupo `IAMAllowedPrincipals` para o recurso. O usuário deve revogar todas as permissões do Lake Formation de `IAMAllowedPrincipals` antes de compartilhar o recurso.

ConcurrentModificationException em grant/revoke solicitações para contas externas

Quando os usuários fazem várias solicitações simultâneas de concessão e and/or revogação de permissão para um diretor nas políticas de tag LF, o Lake Formation lança. ConcurrentModificationException Os usuários precisam capturar a exceção e repetir a falha

na grant/revoke solicitação. Usando versões em lote das operações de `GrantPermissions` `RevokePermissions` /API - BatchGrantPermissionse BatchRevokePermissionsalivia esse problema até certo ponto, reduzindo o número de solicitações simultâneas grant/revoke .

Erro ao usar o Amazon EMR para acessar dados compartilhados por meio de várias contas

Quando você usa o Amazon EMR para acessar dados de outra conta compartilhados com você, algumas bibliotecas do Spark tentarão chamar a operação de API `Glue:GetUserDefinedFunctions`. Como as versões 1 e 2 das permissões AWS RAM gerenciadas não oferecem suporte a essa ação, você recebe a seguinte mensagem de erro:

```
"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"
```

Para resolver esse erro, o administrador do data lake que criou o compartilhamento de recursos deve atualizar as permissões AWS RAM gerenciadas anexadas ao compartilhamento de recursos. A versão 3 das permissões gerenciadas pelo AWS RAM permite que as entidades principais executem a ação `glue:GetUserDefinedFunctions`.

Se você criar um novo compartilhamento de recursos, o Lake Formation aplicará a versão mais recente da permissão AWS RAM gerenciada por padrão, e nenhuma ação será exigida por você. Para habilitar o acesso a dados entre contas para compartilhamentos de recursos existentes, você precisa atualizar as permissões AWS RAM gerenciadas para a versão 3.

Você pode ver as AWS RAM permissões atribuídas aos recursos compartilhados com você em AWS RAM. As permissões incluídas na versão 3 são estas:

```
Databases
  AWSRAMPermissionGlueDatabaseReadWriteForCatalog
  AWSRAMPermissionGlueDatabaseReadWrite

Tables
  AWSRAMPermissionGlueTableReadWriteForCatalog
  AWSRAMPermissionGlueTableReadWriteForDatabase

AllTables
```

```
AWSRAMPermissionGlueAllTablesReadWriteForCatalog
AWSRAMPermissionGlueAllTablesReadWriteForDatabase
```

Para atualizar a versão de permissões AWS RAM gerenciadas dos compartilhamentos de recursos existentes

Você (administrador do data lake) pode atualizar as permissões AWS RAM gerenciadas para uma versão mais recente seguindo as instruções no Guia do AWS RAM usuário ou revogar todas as permissões existentes para o tipo de recurso e concedê-las novamente. Se você revogar as permissões, AWS RAM excluirá o compartilhamento AWS RAM de recursos associado ao tipo de recurso. Quando você concede permissões novamente, AWS RAM cria novos compartilhamentos de recursos anexando a versão mais recente das permissões AWS RAM gerenciadas.

Solução de problemas em esquemas e fluxos de trabalho

Use as informações contidas aqui para ajudar a diagnosticar e corrigir problemas no esquema e no fluxo de trabalho.

Tópicos

- <role-ARN>Meu plano falhou com “Usuário: <user-ARN>não está autorizado a executar: iam: PassRole no recurso:”
- Meu fluxo de trabalho falhou com “Usuário: <user-ARN>não está autorizado a executar: iam: PassRole no recurso:<role-ARN>”
- Um crawler no meu fluxo de trabalho falhou com “o recurso não existe ou o solicitante não está autorizado a acessar as permissões solicitadas”
- Um rastreador no meu fluxo de trabalho falhou com “Ocorreu um erro (AccessDeniedException) ao chamar a CreateTable operação...”

<role-ARN>Meu plano falhou com “Usuário: <user-ARN>não está autorizado a executar: iam: PassRole no recurso:”

Foi feita uma tentativa de criar um esquema por um usuário que não tem permissões suficientes para passar a função escolhida.

Atualize a política do IAM do usuário para poder transmitir a função ou peça que ele escolha uma função diferente com as permissões de senha necessárias.

Para obter mais informações, consulte the section called “Referência de personas e permissões do IAM do Lake Formation”.

Meu fluxo de trabalho falhou com “Usuário: <user-ARN>não está autorizado a executar: iam: PassRole no recurso:<role-ARN>”

A função que você especificou para o fluxo de trabalho não tinha uma política em linha que permitisse que a função se transmitisse sozinha.

Para obter mais informações, consulte the section called “(Opcional) Criar um perfil do IAM para fluxos de trabalho”.

Um crawler no meu fluxo de trabalho falhou com “o recurso não existe ou o solicitante não está autorizado a acessar as permissões solicitadas”

Uma possível causa é que a função passada não tinha permissões suficientes para criar uma tabela no banco de dados de destino. Conceda à função a permissão `CREATE_TABLE` no banco de dados.

Um rastreador no meu fluxo de trabalho falhou com “Ocorreu um erro (AccessDeniedException) ao chamar a CreateTable operação...”

Uma possível causa é que a função do fluxo de trabalho não tinha permissões de localização de dados no local de armazenamento de destino. Conceda permissões de localização de dados para a função.

Para obter mais informações, consulte the section called “`ACESSO À LOCALIZAÇÃO DE DADOS`”.

Problemas conhecidos do AWS Lake Formation

Analise esses problemas conhecidos para AWS Lake Formation.

Tópicos

- Limitação na filtragem de metadados da tabela
- Problema ao renomear uma coluna excluída
- Problema com a exclusão de colunas em tabelas CSV
- As partições da tabela devem ser adicionadas em um caminho comum

- Problema com a criação de um banco de dados durante a criação do fluxo de trabalho
- Problema com a exclusão e a recriação de um usuário
- As operações da API do catálogo de dados não atualizam o valor do parâmetro IsRegisteredWithLakeFormation
- As operações do Lake Formation não oferecem suporte ao AWS Glue Schema Registry

Limitação na filtragem de metadados da tabela

AWS Lake Formation permissões em nível de coluna podem ser usadas para restringir o acesso a colunas específicas em uma tabela. Quando um usuário recupera metadados sobre a tabela usando o console ou uma API como `glue:GetTable`, a lista de colunas no objeto da tabela contém somente os campos aos quais ele tem acesso. É importante entender as limitações dessa filtragem de metadados.

Embora o Lake Formation disponibilize metadados sobre permissões de coluna para serviços integrados, a filtragem real das colunas nas respostas da consulta é de responsabilidade do serviço integrado. Os clientes do Lake Formation que oferecem suporte à filtragem em nível de coluna, incluindo Amazon Athena, Amazon Redshift Spectrum e Amazon EMR, filtram os dados com base nas permissões de coluna registradas no Lake Formation. Os usuários não poderão ler nenhum dado ao qual não devem ter acesso. Atualmente, o ETL AWS Glue não oferece suporte à filtragem de colunas.

> **Note**
>
> Os clusters do EMR não são totalmente gerenciados pela AWS. Portanto, é responsabilidade dos administradores do EMR proteger adequadamente os clusters para evitar o acesso não autorizado aos dados.

Certas aplicações ou formatos podem armazenar metadados adicionais, incluindo nomes e tipos de colunas, no mapa `Parameters` como propriedades da tabela. Essas propriedades são retornadas sem modificações e podem ser acessadas por qualquer usuário com permissão `SELECT` em qualquer coluna.

Por exemplo, o Avro SerDe armazena uma representação JSON do esquema da tabela em uma propriedade de tabela chamada`avro.schema.literal`, que está disponível para todos os usuários com acesso à tabela. Recomendamos que você evite armazenar informações confidenciais

nas propriedades da tabela e esteja ciente de que os usuários podem aprender o esquema completo das tabelas no formato Avro. Essa limitação é específica para os metadados sobre uma tabela.

AWS Lake Formation remove qualquer propriedade da tabela, começando com `spark.sql.sources.schema` ao responder a uma solicitação `glue:GetTable` ou similar, se o chamador não tiver `SELECT` permissões em todas as colunas da tabela. Isso impede que os usuários tenham acesso a metadados adicionais sobre tabelas criadas com o Apache Spark. Quando executadas no Amazon EMR, as aplicações do Apache Spark ainda podem ler essas tabelas, mas certas otimizações podem não ser aplicadas e nomes de colunas com distinção entre maiúsculas e minúsculas não são aceitos. Se o usuário tiver acesso a todas as colunas na tabela, o Lake Formation retornará a tabela sem modificações com todas as propriedades da tabela.

Problema ao renomear uma coluna excluída

Se você usar permissões em nível de coluna para excluir uma coluna e depois renomeá-la, a coluna não será mais excluída das consultas, assim como `SELECT *`.

Problema com a exclusão de colunas em tabelas CSV

Se você criar uma tabela do catálogo de dados com o formato CSV e depois excluir uma coluna do esquema, as consultas poderão retornar dados errados e as permissões em nível de coluna poderão não ser respeitadas.

Solução alternativa: em vez disso, crie uma nova tabela.

As partições da tabela devem ser adicionadas em um caminho comum

O Lake Formation espera que todas as partições de uma tabela estejam em um caminho comum definido no campo de localização da tabela. Quando você usa o crawler para adicionar partições a um catálogo, isso funciona perfeitamente. Mas se você adicionar partições manualmente e essas partições não estiverem no local definido na tabela principal, o acesso aos dados não funcionará.

Problema com a criação de um banco de dados durante a criação do fluxo de trabalho

Ao criar um fluxo de trabalho a partir de um esquema usando o console do Lake Formation, você pode criar o banco de dados de destino, caso ele não exista. Quando você faz isso, o usuário que está conectado recebe a permissão `CREATE_TABLE` no banco de dados criado. No entanto, o

crawler que o fluxo de trabalho gera assume a função do fluxo de trabalho ao tentar criar uma tabela. Isso falha porque a função não possui a permissão `CREATE_TABLE` no banco de dados.

Solução alternativa: se você criar o banco de dados por meio do console durante a configuração do fluxo de trabalho, antes de executar o fluxo de trabalho, deverá conceder à função associada ao fluxo de trabalho a permissão `CREATE_TABLE` no banco de dados que você acabou de criar.

Problema com a exclusão e a recriação de um usuário

O cenário a seguir resulta em permissões errôneas do Lake Formation retornadas por `lakeformation:ListPermissions`:

1. Crie um usuário e conceda permissões do Lake Formation.
2. Exclua o usuário.
3. Recrie o usuário com o mesmo nome.

`ListPermissions` retorna duas entradas, uma para o usuário antigo e outra para o novo usuário. Se você tentar revogar as permissões concedidas ao usuário antigo, as permissões serão revogadas do novo usuário.

As operações da API do catálogo de dados não atualizam o valor do parâmetro **`IsRegisteredWithLakeFormation`**

Há uma limitação conhecida de que as operações da API do catálogo de dados, como `GetTables` e `SearchTables`, não atualizam o valor do parâmetro `IsRegisteredWithLakeFormation` e retornam o padrão, que é falso. É recomendável usar a API `GetTable` para visualizar o valor correto do parâmetro `IsRegisteredWithLakeFormation`.

As operações do Lake Formation não oferecem suporte ao AWS Glue Schema Registry

As operações do Lake Formation não oferecem suporte a AWS Glue tabelas que contenham um `SchemaReference` `StorageDescriptor` para ser utilizado no Registro de Esquemas.

Mensagem de erro atualizada

AWS O Lake Formation atualizou as exceções específicas do recurso para a mensagem de `EntityNotFound` erro geral das seguintes operações de API para atender aos objetivos de segurança e conformidade.

- RevokePermissions
- GrantPermissions
- GetResourceLFTags
- GetTable
- GetDatabase

API AWS Lake Formation

> **Note**
>
> A Referência de API atualizada para o serviço AWS Lake Formation já está disponível.

Sumário

- APIs de permissões
  - Operações
  - Tipos de dados
- Configurações do data lake APIs
  - Operações
  - Tipos de dados
- Integração com o IAM Identity Center APIs
  - Operações
  - Tipos de dados
- APIs de modo de acesso híbrido
  - Operações
  - Tipos de dados
- APIs de fornecimento de credenciais
  - Operações
  - Tipos de dados
- APIs de Tags
  - Operações
  - Tipos de dados
- APIs de filtro de dados
  - Operações
  - Tipos de dados
- Tipos de dados comuns
  - Estrutura ErrorDetail

- Padrões de string

APIs de permissões

A seção API de permissões descreve as operações e os tipos de dados necessários para conceder e revogar permissões no AWS Lake Formation. Consulte o Guia de referência da API do Lake Formation para ver todas as operações e tipos de dados da API do AWS Lake Formation.

Operações

- GrantPermissions
- RevokePermissions
- BatchGrantPermissions
- BatchRevokePermissions
- GetEffectivePermissionsForPath
- ListPermissions
- GetDataLakePrincipal

Tipos de dados

- Recurso
- DatabaseResource
- TableResource
- TableWithColumnsResource
- DataCellsFilterResourcee
- DataLocationResource
- DataLakePrincipal
- PrincipalPermissions
- PrincipalResourcePermissions
- DetailsMap
- ColumnWildcard
- BatchPermissionsRequestEntry

- BatchPermissionsFailureEntry

Configurações do data lake APIs

Esta seção contém as operações da API de configurações do data lake e os tipos de dados para gerenciar os administradores do data lake.

Operações

- GetDataLakeSettings
- PutDataLakeSettings

Tipos de dados

- DataLakeSettings

Integração com o IAM Identity Center APIs

Esta seção contém as operações para criar e gerenciar a integração do Lake Formation com o Centro de Identidade do IAM.

Operações

- CreateLakeFormationIdentityCenterConfiguration
- DeleteLakeFormationIdentityCenterConfiguration
- DescribeLakeFormationIdentityCenterConfiguration
- UpdateLakeFormationIdentityCenterConfiguration

Tipos de dados

- ExternalFilteringConfiguration

APIs de modo de acesso híbrido

A seção API de modo de acesso híbrido descreve as operações e os tipos de dados necessários para configurar o modo de acesso híbrido no AWS Lake Formation. Consulte o Guia de referência da API do Lake Formation para ver todas as operações e tipos de dados da API do AWS Lake Formation.

Operações

- CreateLakeFormationOptIn
- DeleteLakeFormationOptIn
- ListLakeFormationOptIns

Tipos de dados

- Recurso
- DatabaseResource
- TableResource
- Informações sobre o recurso
- LakeFormationOptInsInfo
- DataLocationResource

APIs de fornecimento de credenciais

A seção API de fornecimento de credenciais descreve as operações e os tipos de dados relacionados ao trabalho com o serviço AWS Lake Formation para fornecer credenciais e para registrar e gerenciar um recurso de data lake.

Operações

- RegisterResource
- DeregisterResource
- ListResources
- GetUnfilteredTableMetadata

- GetUnfilteredPartitionsMetadata
- GetTemporaryGluePartitionCredentials
- GetTemporaryGlueTableCredentials
- UpdateResource

Tipos de dados

- FilterCondition
- RowFilter
- ResourceInfo

APIs de Tags

A seção API de tags descreve as operações e os tipos de dados relacionados a uma estratégia de autorização que define um modelo de permissões em atributos ou tags de pares de valores-chave.

Operações

- GetLFTagExpression
- ListLFTagExpressions
- DeleteLFTagExpression
- UpdateLFTagExpression
- CreateLFTagExpression
- AddLFTagsToResource
- RemoveLFTagsFromResource
- GetResourceLFTags
- ListLFTags
- CreateLFTag
- GetLFTag
- UpdateLFTag
- DeleteLFTag
- SearchTablesByLFTags

- SearchDatabasesByLFTags

Tipos de dados

- LFTagKeyResource
- LFTagPolicyResource
- TaggedTable
- TaggedDatabase
- LFTag
- LFTagPair
- LFTagError
- ColumnLFTag

APIs de filtro de dados

As APIs de filtro de dados descrevem como gerenciar filtros de células de dados no AWS Lake Formation.

Operações

- CreateDataCellsFilter
- DeleteDataCellsFilter
- ListDataCellsFilter
- GetDataCellsFilter
- UpdateDataCellsFilter

Tipos de dados

- DataCellsFilter
- RowFilter

Tipos de dados comuns

Tipos de dados comuns descrevem os diversos tipos de dados comuns no AWS Lake Formation.

Estrutura ErrorDetail

Contém detalhes sobre um erro.

Campos

- `ErrorCode` – String UTF-8, superior a 1 e inferior a 255 bytes de comprimento, correspondente a Single-line string pattern.

  O código associado a este erro.
- `ErrorMessage` – string de descrição com no máximo 2.048 bytes de comprimento e correspondente a URI address multi-line string pattern.

  Uma mensagem descrevendo o erro.

Padrões de string

A API usa as seguintes expressões regulares para definir o que é conteúdo válido para vários membros e parâmetros de string:

- Single-line string pattern – `"[\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\t]*"`
- Padrão de string com várias linhas de endereço URI – `"[\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]*"`
- Padrão de string personalizado Nº. 3: `"^\w+\.\w+\.\w+$"`
- Padrão de string personalizado Nº. 4: `"^\w+\.\w+$"`
- Padrão de string personalizado Nº. 5: `"arn:aws:iam::[0-9]*:role/.*"`
- Padrão de string personalizado Nº. 6: `"arn:aws:iam::[0-9]*:user/.*"`
- Padrão de string personalizado Nº. 7: `"arn:aws:iam::[0-9]*:group/.*"`
- Padrão de string personalizado Nº. 8 – `"arn:aws:iam::[0-9]*:saml-provider/.*"`
- Padrão de string personalizado Nº. 9 – `"^([\p{L}\p{Z}\p{N}_.:\/=+\-@%]*)$"`
- Padrão de string personalizado Nº. 10: `“^([\p{L}\p{Z}\p{N}_.:\*\/=+\-@%]*)$”`

- Padrão de string personalizado Nº. 11: “`[\p{L}\p{N}\p{P}]*`”

Regiões aceitas

Esta seção contém informações sobre o suporte Regiões da AWS e a funcionalidade do Lake Formation.

Disponibilidade geral

Para obter o Regiões da AWS suporte de AWS Lake Formation, consulte Lista de AWS serviços disponíveis por região.

Para obter uma lista dos endpoints de serviço do Lake Formation para cada região e as cotas de serviço do Lake Formation, consulte endpoints e cotas do AWS Lake Formation.

AWS GovCloud (US)

Para uma visão geral das diferenças entre AWS GovCloud (US) região e padrão Regiões da AWS, consulte Como AWS Lake Formation difere para AWS GovCloud (US).

Otimização de transações e armazenamento

As tabelas controladas, o suporte a transações e os recursos de otimização de armazenamento do Lake Formation estão disponíveis no seguinte: Regiões da AWS

Nome da região	Parâmetro da região	Endpoint
Leste dos EUA (Norte da Virgínia)	us-east-1	lakeformation.us-east-1.amazonaws.com lakeformation-fips.us-east-1.amazonaws.com
Leste dos EUA (Ohio)	us-east-2	lakeformation.us-east-2.amazonaws.com lakeformation-fips.us-east-2.amazonaws.com

Nome da região	Parâmetro da região	Endpoint
Oeste dos EUA (Oregon)	us-west-2	lakeformation.us-west-2.amazonaws.com lakeformation-fips.us-west-2.amazonaws.com
Ásia-Pacífico (Mumbai)	ap-south-1	lakeformation.ap-south-1.amazonaws.com
Ásia-Pacífico (Seul)	ap-northeast-2	lakeformation.ap-northeast-2.amazonaws.com
Ásia-Pacífico (Singapura)	ap-southeast-1	lakeformation.ap-southeast-1.amazonaws.com
Ásia-Pacífico (Sydney)	ap-southeast-2	lakeformation.ap-southeast-2.amazonaws.com
Ásia-Pacífico (Tóquio)	ap-northeast-1	lakeformation.ap-northeast-1.amazonaws.com
Europa (Frankfurt)	eu-central-1	lakeformation.eu-central-1.amazonaws.com
Europa (Irlanda)	eu-west-1	lakeformation.eu-west-1.amazonaws.com
Europa (Londres)	eu-west-2	lakeformation.eu-west-2.amazonaws.com
Europa (Estocolmo)	eu-north-1	lakeformation.eu-north-1.amazonaws.com

Nome da região	Parâmetro da região	Endpoint
Canadá (Central)	ca-central-1	lakeformation.ca-central-1.amazonaws.com
América do Sul (São Paulo)	sa-east-1	lakeformation.sa-east-1.amazonaws.com

Histórico do documento para AWS Lake Formation

A tabela a seguir descreve mudanças importantes na documentação do AWS Lake Formation.

Alteração	Descrição	Data
Documentação atualizada para controle de acesso baseado em atributos	O Lake Formation permite que você conceda permissões às entidades principais com atributos correspondentes nos recursos do Data Catalog. Para acessar mais informações, consulte Controle de acesso por atributo.	17 de abril de 2025
Integração de tabelas do Amazon S3 com e AWS Lake FormationAWS Glue Data Catalog	Agora você pode integrar e catalogar tabelas do S3 como AWS Glue Data Catalog objetos e registrar o catálogo como um local de dados do Lake Formation. Para acessar mais informações, consulte Criar um catálogo de Tabelas do Amazon S3 no AWS Glue Data Catalog.	13 de março de 2025
`AWSLakeFormationCrossAccountManager` Política atualizada do Lake Formation.	O Lake Formation aprimorou a AWSLakeFormationCrossAccountManagerpolítica substituindo o operador de `StringLike` condição pelo `ArnLike` operador que permite ao IAM realizar a verificação do formato ARN.	25 de janeiro de 2025

Atualização de alterações em políticas	Documentou a mudança na AWSLakeFormationDataAdminpolítica.	3 de dezembro de 2024
Atualizações de vários catálogos	AWS Glue Data Catalog permite criar catálogos federados e unificar dados entre data lakes do Amazon S3 e armazéns de dados do Amazon Redshift, bem como integrar dados de seus bancos de dados operacionais, como o Amazon DynamoDB, e de fontes de dados de terceiros , como Snowflake, MySQL e assim por diante. Para acessar mais informaçõ es, consulte Trazer seus dados para o AWS Glue Data Catalog.	3 de dezembro de 2024
Atualização da documentação para expressões de tag LF	Você pode salvar expressões de tag LF e reutilizá-las para conceder permissões nos recursos do Data Catalog. Para acessar mais informaçõ es, consulte Gerenciar expressões de tag LF.	7 de novembro de 2024
Documentação atualizad a sobre visualizações do Catálogo de Dados	Você pode criar visualiza ções AWS Glue Data Catalog usando, AWS Glue APIs além de usar o Amazon Athena e o Amazon Redshift. DDLs Para ter mais informações, consulte Criar visualizações do Catálogo de Dados.	7 de agosto de 2024

Documentação adicionada sobre o fornecimento de credenciais auditáveis	O Lake Formation permite que você inclua o contexto do usuário do IAM Identity Center nos CloudTrail eventos e, em seguida, acompanhe os usuários que acessam seus recursos. Para obter mais informações, consulte Incluir o contexto do usuário do IAM Identity Center nos CloudTrail registros.	14 de julho de 2024
Atualização de alterações em políticas	Documentou a alteração (declaração adicionada IDs e permissões redundantes removidas) nas políticas AWSLakeFormationCrossAccountManager e. AWSLake Formation DataAdmin	14 de março de 2024
Configuração atualizada para o Lake Formation	As etapas na seção Set up AWS Lake Formation foram atualizadas.	7 de fevereiro de 2024
Atualização de alterações em políticas	Adicionadas novas permissões à política em linha do perfil vinculado ao serviço. Para ter mais informações, consulte Using service-linked roles for Lake Formation.	7 de fevereiro de 2024
Atualização de alterações em políticas	Documentou a mudança na LakeFormationDataAccessServiceRolePolicypolítica.	2 de fevereiro de 2024

Limitações consolidadas do Lake Formation	Foi criada uma seção unificada das limitações e considerações do Lake Formation. Para obter mais informações, consulte Lake Formation limitations.	15 de dezembro de 2023
Foi adicionada documentação sobre compactação do Iceberg.	Para melhor desempenho de leitura por serviços de AWS análise, como Athena e Amazon EMR, e trabalhos de AWS Glue ETL, AWS Glue Data Catalog fornece compactação gerenciada (um processo que compacta pequenos objetos do Amazon S3 em objetos maiores) para tabelas Iceberg no catálogo de dados. Para obter mais informações, consulte Como otimizar tabelas do Iceberg.	25 de novembro de 2023
Foi adicionada documentação sobre a integração com o Centro de Identidade do IAM.	As integrações com o Centro de Identidade do IAM permitem que usuários e grupos acessem os recursos do catálogo de dados aplicando as permissões do Lake Formation. Para obter mais informações, consulte IAM Identity Center integration.	25 de novembro de 2023

Foi adicionada documentação sobre visualizações do catálogo de dados.	Você pode criar visualizações AWS Glue Data Catalog que façam referência a até 10 tabelas usando editores SQL Amazon Athena ou Amazon Redshift. Para obter mais informações, consulte Creating views.	25 de novembro de 2023
Atualização da mudança de política	Documentou a mudança na AWSLakeFormationCrossAccountManagerpolítica.	25 de outubro de 2023
Adicionada documentação do modo de acesso híbrido	O modo de acesso híbrido oferece a flexibilidade de habilitar seletivamente as permissões do Lake Formation para bancos de dados e tabelas no seu AWS Glue Data Catalog. Com o modo de acesso híbrido, agora você tem um caminho incremental que permite definir permissões do Lake Formation para um conjunto específico de usuários sem interromper as políticas de permissão de outros usuários ou workload existentes. Para obter mais informações, consulte Modo de acesso híbrido.	26 de setembro de 2023

Adicionada documentação para a criação de tabelas do Apache Iceberg	Agora você pode criar tabelas do Apache Iceberg que usam o formato de dados Apache Parquet AWS Glue Data Catalog com dados residentes no Amazon S3. Para obter mais informações, consulte Como criar tabelas do Apache Iceberg.	16 de agosto de 2023
Adicionada documentação para acesso a dados entre regiões	O Lake Formation suporta a consulta de tabelas do Catálogo de Dados em todas as AWS regiões. Você pode acessar dados em uma região de outras regiões usando Athena, Amazon EMR e executar AWS Glue ETL criando links de recursos em outras regiões apontando para os bancos de dados e tabelas de origem. Você pode conectar o catálogo de dados a repositórios externos que armazenam metadados para seus dados do Amazon S3 e gerenciar com segurança as permissões de acesso aos dados usando o AWS Lake Formation. Para obter mais informações, consulte Acesso a tabelas entre regiões.	30 de junho de 2023

Reorganização do conteúdo	Capítulos reorganizados no guia para acompanhar a jornada do usuário de Lake Formation.	15 de maio de 2023
Adicionada a documentação da federação do HMS	Você pode conectar o catálogo de dados a repositórios externos que armazenam metadados para seus dados do Amazon S3 e gerenciar com segurança as permissões de acesso aos dados usando o AWS Lake Formation. Para obter mais informações, consulte Gerenciamento de permissões em conjuntos de dados que usam repositórios externos.	15 de abril de 2023
Adicionada a documentação da unidade de compartilhamento de dados do Amazon Redshift	Agora você pode gerenciar dados com segurança em uma unidade de compartilhamento de dados do Amazon Redshift usando as permissões do Lake Formation. O Lake Formation oferece suporte ao licenciamento de acesso aos seus dados por meio AWS Data Exchange de. Para obter mais informações, consulte Compartilhamento de dados em AWS Lake Formation.	30 de novembro de 2022

Suporte para compartilhamento de dados entre contas diretamente com as entidades principais	Adicionadas informações sobre a unidade de compartilhamento de dados diretamente com as entidades principais do IAM em outra conta. Para obter mais informações, consulte Compartilhamento de dados entre contas no AWS Lake Formation.	10 de novembro de 2022
Support para compartilhamento de dados AWS RAM habilitado usando TBAC	Foram adicionadas informações sobre o método LF-TBAC de conceder o uso de permissões do Catálogo de Dados para concessões entre contas. AWS Resource Access Manager	10 de novembro de 2022
Adicionada uma seção sobre como trabalhar com outros serviços	Foram adicionadas informações sobre como AWS serviços como Athena, AWS Glue Redshift Spectrum e Amazon EMR podem usar o Lake Formation para acessar com segurança dados em locais do Amazon S3 registrados no Lake Formation. Para obter mais informações, consulte Trabalhando com outros AWS serviços.	10 de novembro de 2022

???	Adicionadas informações sobre como solucionar um erro ao usar o Amazon EMR para acessar dados entre contas. Para obter mais informações, consulte Erro ao usar o Amazon EMR para acessar dados compartilhados por meio de várias contas.	7 de novembro de 2022
Atualizações no compartilhamento de recursos entre contas	Adicionada uma descrição de como os compartilhamentos de recursos entre contas funcionam no Lake Formation. Documentou a mudança na AWSLakeFormationCrossAccountManagerpolítica.	6 de maio de 2022
Novos tutoriais	Adicionados novos tutoriais para criar tabelas controladas, proteger data lakes e compartilhar data lakes. Para obter mais informações, consulte a seção Começar.	20 de abril de 2022
Nova página inicial do Lake Formation	A página inicial do Lake Formation foi atualizada para incluir links para tutoriais que fornecem step-by-step instruções sobre como criar um data lake, ingerir dados, compartilhar e proteger data lakes usando o Lake Formation.	20 de abril de 2022

Suporte para fornecimento de credenciais	Adicionadas informações sobre o fornecimento de credenciais, que oferece suporte ao Lake Formation para permitir que serviços de terceiros integrem-se ao Lake Formation usando operações de API de fornecimento automático de credenciais. Para obter mais informaçõ es, consulte Como funciona o fornecimento de credenciais no Lake Formation.	28 de fevereiro de 2022
Suporte para tabelas controlad as e filtragem avançada de dados	Foram adicionadas informaçõ es sobre tabelas governada s, que suportam transações ACID, compactação automátic a de dados e consultas de viagem no tempo. Adicionadas informações sobre a criação de filtros de dados oferecem suporte à segurança por coluna, segurança por linha e segurança por célula. Para obter mais informações, consulte Tabelas governadas no Lake Formation e Filtragem de dados e segurança por célula no Lake Formation.	30 de novembro de 2021

Suporte para endpoints da interface da VPC	Foram adicionadas informações sobre a criação de um endpoint de interface de nuvem privada virtual (VPC) para o Lake Formation, para que a comunicação entre sua VPC e o Lake Formation seja conduzida de forma completa e segura na rede. AWS Para obter mais informações, consulte Como usar o Lake Formation com endpoints da VPC.	11 de outubro de 2021
Suporte a políticas de endpoint da VPC	Foram adicionadas informações sobre o suporte a políticas de endpoint do nuvem privada virtual (VPC) em Lake Formation. Para obter mais informações, consulte Como usar o Lake Formation com endpoints da VPC.	11 de outubro de 2021
Suporte para controle de acesso baseado em tags	O controle de acesso baseado em tags do Lake Formation fornece uma maneira nova e mais escalável de gerenciar o acesso aos recursos do catálogo de dados e aos dados subjacentes usando tags do LF. Para obter mais informações, consulte Controle de acesso baseado em tags do Lake Formation.	7 de maio de 2021

Novo requisito de aceitação para filtragem de dados no Amazon EMR.	Adicionadas informações sobre a exigência de se cadastrar para permitir que o Amazon EMR filtre dados gerenciados pelo Lake Formation. Para obter mais informações, consulte Permitir filtragem de dados no Amazon EMR.	9 de outubro de 2020
Suporte para conceder permissões completas entre contas nos bancos de dados do catálogo de dados	Adicionadas informações sobre a concessão de permissões completas do Lake Formation nos bancos de dados do catálogo de dados em todas as contas da AWS , incluindo `CREATE_TABLE` . Para obter mais informações, consulte Compartilhamento de bancos de dados do Data Catalog.	1.º de outubro de 2020
Support para autenticação de Amazon Athena usuários por meio do SAML.	Adicionadas informações sobre o suporte para usuários do Athena que se conectam por meio do driver JDBC ou ODBC, e se autenticam por meio de provedores de identidade SAML, como Okta e Microsoft Active Directory Federation Service (AD FS). Para obter mais informações, consulte Integrações do serviço AWS com o Lake Formation.	30 de setembro de 2020

Suporte para acesso entre contas com um catálogo de dados criptografado	Adicionadas informaçõ es sobre a concessão de permissões entre contas quando o catálogo de dados é criptografado. Para obter mais informações, consulte Requisitos para Acesso Entre Contas.	30 de julho de 2020
Suporte para acesso entre contas ao data lake	Foram adicionadas informaçõ es sobre a concessão de AWS Lake Formation permissões em bancos de dados e tabelas do Catálogo de Dados para AWS contas e organizaç ões externas e sobre como acessar objetos do Catálogo de Dados compartilhados de contas externas. Para obter mais informações, consulte Acesso entre contas.	7 de julho de 2020
Integração com o Amazon Quick	Foram adicionadas informaçõ es sobre como conceder permissões do Lake Formation aos usuários do Quick Enterprise Edition para que eles possam acessar conjuntos de dados residente s em locais registrados do Amazon S3. Para obter mais informações, consulte Concessão de permissões do catálogo de dados.	29 de junho de 2020

Atualizações nos capítulos de configuração e introdução	Reorganizou e aprimorou os capítulos Configuração e Introdução. Atualizou as permissões recomendadas AWS Identity and Access Management (IAM) para o administrador do data lake.	27 de fevereiro de 2020
Support for AWS Key Management Service	Foram adicionadas informaçõ es sobre como o suporte do Lake Formation para AWS Key Managemen t Service (AWS KMS) simplifica a configuração de serviços integrados para ler e gravar dados criptogra fados em locais registrados do Amazon Simple Storage Service (Amazon S3). Foram adicionadas informações sobre como registrar locais do Amazon S3 que são criptogra fados com. AWS KMS keys Para obter mais informações, consulte the section called “Adicionar uma localização do Amazon S3 ao seu data lake”.	27 de fevereiro de 2020
Atualizações nos esquemas e nas políticas do IAM do administrador do data lake	Parâmetros de entrada esclarecidos para esquemas de banco de dados increment ais. Atualizou as políticas do IAM necessárias para um administrador de data lake.	20 de dezembro de 2019

Reescrito o capítulo de segurança e revisões do capítulo de atualização	Melhoria dos capítulos de segurança e atualização.	29 de outubro de 2019
A super permissão substitui todas as permissões	Os capítulos de Segurança e Atualização foram atualizados para refletir a substituição da permissão `All` por `Super`.	10 de outubro de 2019
Adições, correções e esclarecimentos	Foram feitas adições, correções e esclarecimentos com base nas opiniões. Revisado o capítulo sobre segurança. Os capítulos de Segurança e Atualização foram atualizados para refletir a substituição do grupo `Everyone` por `IAMAllowedPrincipals` .	11 de setembro de 2019
Novo guia	Este é o lançamento inicial do Guia do desenvolvedor do AWS Lake Formation .	8 de agosto de 2019

Glossário da AWS

Para obter a terminologia mais recente da AWS, consulte o glossário da AWS na Referência do Glossário da AWS.

As traduções são geradas por tradução automática. Em caso de conflito entre o conteúdo da tradução e da versão original em inglês, a versão em inglês prevalecerá.