Guida per gli sviluppatori

AWS Lake Formation

AWS Lake Formation: Guida per gli sviluppatori

Table of Contents

Cos'è AWS Lake Formation?

Benvenuto nella guida per sviluppatori di AWS Lake Formation.

AWS Lake Formationti aiuta a governare, proteggere e condividere a livello globale i dati per l'analisi e l'apprendimento automatico. Con Lake Formation, puoi gestire un controllo granulare degli accessi per i dati del tuo data lake su Amazon Simple Storage Service (Amazon S3) e i relativi metadati. AWS Glue Data Catalog

Lake Formation fornisce il proprio modello di autorizzazioni che amplia il modello di autorizzazioni IAM. Il modello di autorizzazioni di Lake Formation consente un accesso granulare ai dati archiviati nei data lake attraverso un semplice meccanismo di concessione o revoca, proprio come un sistema di gestione di database relazionali (RDBMS). Le autorizzazioni di Lake Formation vengono applicate utilizzando controlli granulari a livello di colonna, riga e cella nei servizi di AWS analisi e apprendimento automatico, tra cui Amazon Athena, Amazon Amazon QuickSight Redshift Spectrum, Amazon EMR e. AWS Glue

La modalità di accesso ibrido di Lake Formation ti AWS Glue Data Catalog consente di proteggere e accedere ai dati catalogati utilizzando sia le autorizzazioni di Lake Formation che le politiche di autorizzazione IAM per Amazon S3 e azioni. AWS Glue Con la modalità di accesso ibrida, gli amministratori dei dati possono integrare le autorizzazioni di Lake Formation in modo selettivo e incrementale, concentrandosi su un caso d'uso del data lake alla volta.

Lake Formation consente inoltre di condividere i dati internamente ed esternamente tra più AWS organizzazioni o direttamente con i responsabili IAM in un altro accountAccount AWS, fornendo un accesso granulare ai metadati e ai dati sottostanti. AWS Glue Data Catalog

Argomenti

Caratteristiche di Lake Formation

Lake Formation ti aiuta a scomporre i silos di dati e a combinare diversi tipi di dati strutturati e non strutturati in un repository centralizzato. Innanzitutto, identifica gli archivi di dati esistenti in Amazon S3 o nei database relazionali e NoSQL e sposta i dati nel tuo data lake. Quindi scansiona, cataloga e prepara i dati per l'analisi. Successivamente, offri ai tuoi utenti un accesso self-service sicuro ai dati tramite i servizi di analisi di loro scelta.

Argomenti

Inserimento e gestione dei dati

Importa dati da database già presenti AWS

Dopo aver specificato dove si trovano i database esistenti e fornito le credenziali di accesso, Lake Formation legge i dati e i relativi metadati (schema) per comprendere il contenuto della fonte di dati. Quindi importa i dati nel nuovo data lake e registra i metadati in un catalogo centrale. Con Lake Formation, puoi importare dati da database MySQL, PostgreSQL, SQL Server, MariaDB e Oracle in esecuzione in Amazon RDS o ospitati in Amazon EC2. Sono supportati sia il caricamento di dati in blocco che quello incrementale.

Importa dati da altre fonti esterne

Puoi utilizzare Lake Formation per spostare i dati dai database locali connettendoti a Java Database Connectivity (JDBC). Identifica le fonti di destinazione e fornisci le credenziali di accesso nella console e Lake Formation legge e carica i tuoi dati nel data lake. Per importare dati da database diversi da quelli sopra elencati, puoi creare lavori ETL personalizzati con. AWS Glue

Cataloga ed etichetta i tuoi dati

Puoi usare AWS Glue i crawler per leggere i tuoi dati in Amazon S3 ed estrarre schemi di database e tabelle e archiviare tali dati in un file ricercabile. AWS Glue Data Catalog Quindi, usa Lake Formation Controllo degli accessi basato su tag Lake Formation (TBAC) per gestire le autorizzazioni su database, tabelle e colonne. Per ulteriori informazioni sull'aggiunta di tabelle al Data Catalog, consulta. Creazione di tabelle e database del catalogo dati

Gestione della sicurezza

Definisci e gestisci i controlli di accesso

Lake Formation offre un unico posto per gestire i controlli di accesso per i dati nel tuo data lake. È possibile definire politiche di sicurezza che limitano l'accesso ai dati a livello di database, tabella, colonna, riga e cella. Queste policy si applicano agli utenti e ai ruoli IAM e agli utenti e ai gruppi durante la federazione tramite un provider di identità esterno. Puoi utilizzare controlli granulari per accedere ai dati protetti da Lake Formation all'interno di Amazon Redshift Spectrum, Athena, ETL AWS Glue e Amazon EMR per Apache Spark. Ogni volta che crei identità IAM, assicurati di seguire le best practice IAM. Per ulteriori informazioni, consulta le best practice di sicurezza nella Guida per l'utente IAM.

Modalità di accesso ibrida

La modalità di accesso ibrido di Lake Formation offre la flessibilità necessaria per abilitare selettivamente le autorizzazioni di Lake Formation per database e tabelle del tuo. AWS Glue Data Catalog Con la modalità di accesso ibrido, ora disponi di un percorso incrementale che ti consente di impostare le autorizzazioni di Lake Formation per un set specifico di utenti senza interrompere le politiche di autorizzazione di altri utenti o carichi di lavoro esistenti. Per ulteriori informazioni, consulta Modalità di accesso ibrida.

Implementa la registrazione degli audit

Lake Formation fornisce registri di controllo completi CloudTrail per monitorare l'accesso e dimostrare la conformità con le politiche definite a livello centrale. Puoi controllare la cronologia di accesso ai dati attraverso i servizi di analisi e machine learning che leggono i dati nel tuo data lake tramite Lake Formation. In questo modo puoi vedere quali utenti o ruoli hanno tentato di accedere a quali dati, con quali servizi e quando. Puoi accedere ai log di controllo nello stesso modo in cui accedi a qualsiasi altro CloudTrail registro utilizzando le CloudTrail API e la console. Per ulteriori informazioni sui CloudTrail log, consulta. Registrazione di logAWSChiamate API Lake Formation utilizzandoAWS CloudTrail

Sicurezza a livello di riga e cella

Lake Formation fornisce filtri di dati che consentono di limitare l'accesso a una combinazione di colonne e righe. Utilizza la sicurezza a livello di riga e cella per proteggere i dati sensibili come le informazioni personali identificabili (PII). Per ulteriori informazioni sulla sicurezza a livello di riga, consulta. Panoramica del filtraggio dei dati

Controllo degli accessi basato su tag

Usa il controllo degli accessi basato su tag Lake Formation per gestire centinaia o addirittura migliaia di autorizzazioni per i dati creando etichette personalizzate chiamate LF-Tags. Ora puoi definire i tag LF e allegarli a database, tabelle o colonne. Quindi, condividi l'accesso controllato tra i servizi di analisi, machine learning (ML) ed estrazione, trasformazione e caricamento (ETL) per il consumo. I tag LF assicurano che la governance dei dati possa essere scalata facilmente sostituendo le definizioni delle politiche di migliaia di risorse con alcuni tag logici. Lake Formation fornisce una ricerca testuale su questi metadati, in modo che gli utenti possano trovare rapidamente i dati che devono analizzare.

Accesso tra account

Le funzionalità di gestione delle autorizzazioni di Lake Formation semplificano la protezione e la gestione dei data lake distribuiti su più AWS account attraverso un approccio centralizzato, fornendo un controllo granulare degli accessi al Data Catalog e alle sedi Amazon S3. Per ulteriori informazioni, consulta Condivisione dei dati tra account in Lake Formation.

Condivisione dei dati

La funzionalità di condivisione dei dati consente di configurare le autorizzazioni per set di dati archiviati in diverse fonti di dati come Amazon Redshift senza migrare dati o metadati in Amazon S3 o. AWS Glue Data Catalog Puoi utilizzare i seguenti metodi per condividere dati in Lake Formation:

Per ulteriori informazioni, consulta Condivisione dei dati in Lake Formation.

- Integrazione di Lake Formation con la condivisione dei dati di Amazon Redshift: utilizza Lake Formation per gestire centralmente le autorizzazioni di accesso a livello di database, tabelle, colonne e righe delle condivisioni di dati Amazon Redshift e limitare l'accesso degli utenti agli oggetti all'interno di un datashare.
- Connessione AWS Glue Data Catalog a metastore esterni: connettiti AWS Glue Data Catalog a metastore esterni per gestire le autorizzazioni di accesso ai set di dati in Amazon S3 utilizzando Lake Formation. Non è necessaria alcuna migrazione dei metadati in. AWS Glue Data Catalog

  Per ulteriori informazioni, consultare Gestione delle autorizzazioni sui set di dati che utilizzano metastore esterni
- Integrazione di Lake Formation con AWS Data Exchange — Lake Formation supporta la concessione di licenze di accesso ai dati tramite. AWS Data Exchange Se sei interessato a

concedere in licenza i tuoi dati di Lake Formation, consulta Cosa c'è AWS Data Exchange nella Guida per l'AWS Data Exchangeutente.

AWS Lake Formation: Come funziona

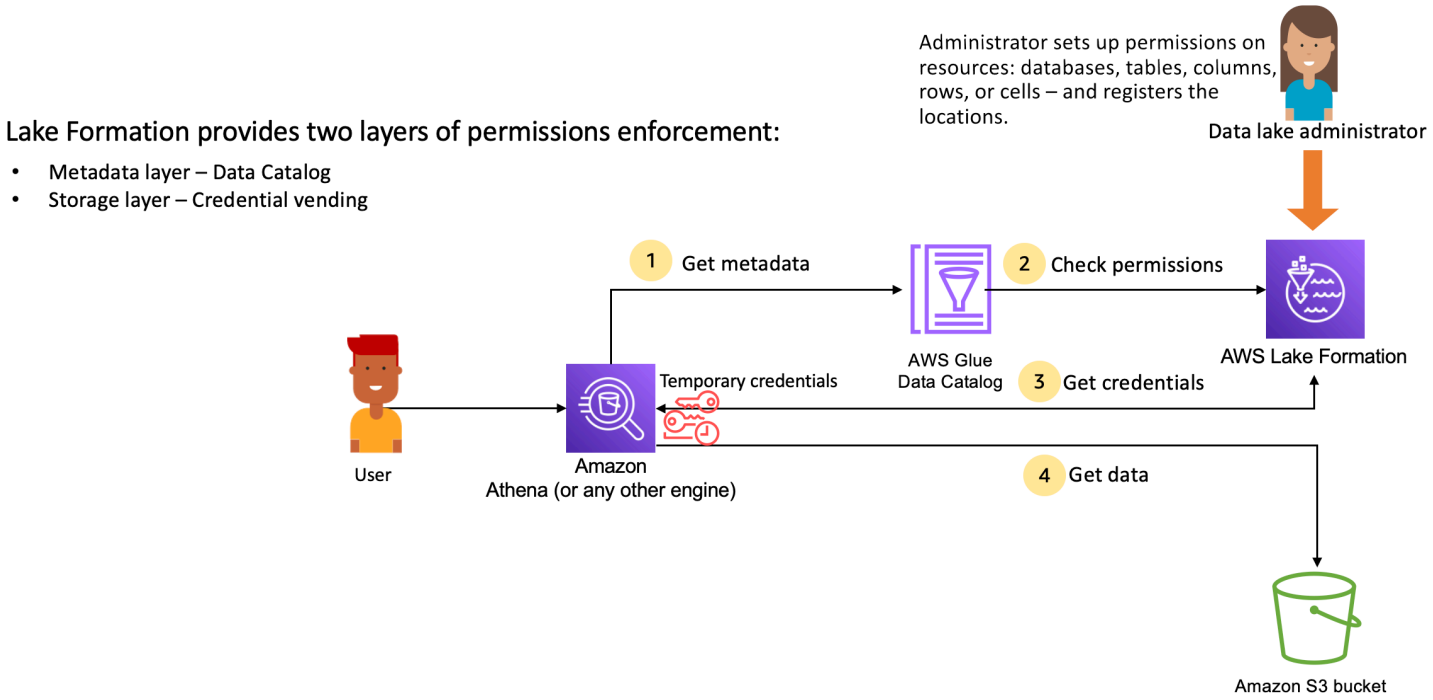
AWS Lake Formationfornisce un modello di autorizzazioni del sistema di gestione dei database relazionali (RDBMS) per concedere o revocare l'accesso alle risorse del Data Catalog come database, tabelle e colonne con dati sottostanti in Amazon S3. Le autorizzazioni Lake Formation, facili da gestire, sostituiscono le complesse policy dei bucket di Amazon S3 e le politiche IAM corrispondenti.

In Lake Formation, puoi implementare le autorizzazioni su due livelli:

- Applicazione delle autorizzazioni a livello di metadati sulle risorse del Data Catalog come database e tabelle
- Gestione delle autorizzazioni di accesso allo storage sui dati sottostanti archiviati in Amazon S3 per conto di motori integrati

Flusso di lavoro per la gestione delle autorizzazioni di Lake Formation

Lake Formation si integra con i motori analitici per interrogare gli archivi di dati e gli oggetti di metadati di Amazon S3 registrati con Lake Formation. Il diagramma seguente illustra come funziona la gestione delle autorizzazioni in Lake Formation.



Fasi di alto livello per la gestione dei permessi di Lake Formation

Prima che Lake Formation possa fornire controlli di accesso per i dati nel tuo data lake, un amministratore del data lake o un utente con autorizzazioni amministrative imposta le politiche utente delle singole tabelle Data Catalog per consentire o negare l'accesso alle tabelle Data Catalog utilizzando le autorizzazioni di Lake Formation.

Quindi, l'amministratore del data lake o un utente delegato dall'amministratore concede le autorizzazioni di Lake Formation agli utenti sui database e sulle tabelle di Data Catalog e registra la posizione Amazon S3 della tabella con Lake Formation.

1. Ottieni metadati: un principale (utente) invia una query o uno script ETL a un motore di analisi integrato come Amazon Athena, Amazon EMR o Amazon AWS Glue Redshift Spectrum. Il motore analitico integrato identifica la tabella richiesta e invia una richiesta di metadati al Data Catalog.
2. Controlla le autorizzazioni: il Data Catalog controlla le autorizzazioni dell'utente con Lake Formation e, se l'utente è autorizzato ad accedere alla tabella, restituisce al motore i metadati che l'utente può vedere.
3. Ottieni credenziali: il Data Catalog consente al motore di sapere se la tabella è gestita da Lake Formation o meno. Se i dati sottostanti sono registrati con Lake Formation, il motore analitico richiede a Lake Formation di fornire l'accesso ai dati concedendo un accesso temporaneo.
4. Ottieni dati: se l'utente è autorizzato ad accedere alla tabella, Lake Formation fornisce l'accesso temporaneo al motore analitico integrato. Utilizzando l'accesso temporaneo, il motore analitico

recupera i dati da Amazon S3 ed esegue i filtri necessari come il filtraggio di colonne, righe o celle. Quando il motore termina l'esecuzione del lavoro, restituisce i risultati all'utente. Questo processo è chiamato vendita di credenziali.

Se la tabella non è gestita da Lake Formation, la seconda chiamata dal motore di analisi viene effettuata direttamente ad Amazon S3. La policy del bucket Amazon S3 interessata e la politica utente IAM vengono valutate per l'accesso ai dati.

Ogni volta che si utilizzano le policy IAM, assicurati di seguire le best practice IAM. Per ulteriori informazioni, consulta Best Practice di sicurezza in IAM nella Guida per l'utente di IAM.

Argomenti

Autorizzazioni per i metadati

Lake Formation fornisce l'autorizzazione e il controllo degli accessi per il Data Catalog. Quando un ruolo IAM effettua una chiamata all'API Data Catalog da qualsiasi sistema, Data Catalog verifica le autorizzazioni relative ai dati dell'utente e restituisce solo i metadati a cui l'utente dispone delle autorizzazioni di accesso. Ad esempio, se un ruolo IAM ha accesso a una sola tabella all'interno di un database e un servizio o un utente che assume il ruolo esegue l'`GetTables`operazione, la risposta conterrà solo una tabella, indipendentemente dal numero di tabelle nel database.

Impostazioni predefinite: autorizzazioni **`IAMAllowedPrincipal`** di gruppo

AWS Lake Formation, per impostazione predefinita, imposta le autorizzazioni per tutti i database e le tabelle su un gruppo virtuale denominato. `IAMAllowedPrincipal` Questo gruppo è unico e visibile solo all'interno di Lake Formation. Il `IAMAllowedPrincipal` gruppo include tutti i responsabili IAM che hanno accesso alle risorse di Data Catalog tramite le politiche principali e le politiche AWS Glue delle risorse IAM. Se queste autorizzazioni esistono su un database o una tabella, a tutti i principali verrà concesso l'accesso al database o alla tabella.

Se desideri fornire autorizzazioni più granulari su un database o una tabella, rimuovi `IAMAllowedPrincipal` l'autorizzazione e Lake Formation applica tutte le altre politiche associate a quel database o tabella. Ad esempio, se esiste una politica che consente all'utente A

di accedere al database A con `DESCRIBE` le autorizzazioni ed `IAMAllowedPrincipal` esiste con tutte le autorizzazioni, l'utente A continuerà a eseguire tutte le altre azioni, fino alla revoca dell'autorizzazione. `IAMAllowedPrincipal`

Inoltre, per impostazione predefinita, il `IAMAllowedPrincipal` gruppo dispone delle autorizzazioni su tutti i nuovi database e tabelle al momento della creazione. Esistono due configurazioni che controllano questo comportamento. La prima è a livello di account e regione, il che consente questa operazione per i database appena creati, mentre la seconda è a livello di database. Per modificare l'impostazione predefinita, vedere. Modifica il modello di autorizzazione predefinito o utilizza la modalità di accesso ibrida

Concessione di autorizzazioni

Gli amministratori di Data Lake possono concedere le autorizzazioni di Data Catalog ai responsabili in modo che i responsabili possano creare e gestire database e tabelle e possano accedere ai dati sottostanti.

Autorizzazioni a livello di database e tabella

Quando concedi le autorizzazioni all'interno di Lake Formation, il concedente deve specificare il principale a cui concedere le autorizzazioni, le risorse a cui concedere le autorizzazioni e le azioni che il beneficiario deve avere accesso a eseguire. Per la maggior parte delle risorse all'interno di Lake Formation, l'elenco principale e le risorse per concedere le autorizzazioni sono simili, ma le azioni che un beneficiario può eseguire variano in base al tipo di risorsa. Ad esempio, `SELECT` le autorizzazioni sono disponibili per le tabelle per leggere le tabelle, ma le `SELECT` autorizzazioni non sono consentite per i database. L'`CREATE_TABLE`autorizzazione è consentita per i database, ma non per le tabelle.

È possibile concedere AWS Lake Formation le autorizzazioni utilizzando due metodi:

- Metodo di risorsa denominato: consente di scegliere i nomi di database e tabelle concedendo le autorizzazioni agli utenti.
- Controllo degli accessi basato su tag LF (LF-TBAC): gli utenti creano tag LF, li associano alle risorse del catalogo dati, concedono l'`Describe`autorizzazione sui tag LF, associano le autorizzazioni ai singoli utenti e scrivono politiche di autorizzazione LF utilizzando i tag LF a diversi utenti. Tali politiche basate su LF-Tag si applicano a tutte le risorse del Data Catalog associate a tali valori LF-Tag.

> **Note**
>
> I tag LF sono esclusivi di Lake Formation. Sono visibili solo in Lake Formation e non devono essere confusi con i tag AWS delle risorse.

LF-TBAC è una funzionalità che consente agli utenti di raggruppare le risorse in categorie di tag LF definite dall'utente e di applicare le autorizzazioni a tali gruppi di risorse. Pertanto, è il modo migliore per scalare le autorizzazioni su un numero enorme di risorse del Data Catalog.

Per ulteriori informazioni, consulta Controllo degli accessi basato su tag Lake Formation.

Quando concedi le autorizzazioni a un responsabile, Lake Formation valuta le autorizzazioni come un'unione di tutte le politiche per quell'utente. Ad esempio, se si dispone di due criteri in una tabella per un principale in cui un criterio concede le autorizzazioni alle colonne col1, col2 e col3 tramite il metodo di risorsa denominato e l'altro criterio concede le autorizzazioni alla stessa tabella e principale a col5 e col6 tramite LF-tags, le autorizzazioni effettive saranno un'unione delle autorizzazioni che sarebbero col1, col2, col3, col5 e col6. Ciò include anche i filtri e le righe di dati.

Autorizzazioni per la localizzazione dei dati

Le autorizzazioni di localizzazione dei dati offrono agli utenti non amministrativi la possibilità di creare database e tabelle in posizioni Amazon S3 specifiche. Se un utente tenta di creare un database o una tabella in una posizione per cui non dispone delle autorizzazioni necessarie, l'operazione di creazione ha esito negativo. Questo serve a impedire agli utenti di creare tabelle in posizioni arbitrarie all'interno del data lake e consente di controllare dove tali utenti possono leggere e scrivere i dati. Esiste un'autorizzazione implicita durante la creazione di tabelle nella posizione Amazon S3 all'interno del database in cui viene creata. Per ulteriori informazioni, consulta Concessione delle autorizzazioni per la localizzazione dei dati.

Crea autorizzazioni per tabelle e database

Per impostazione predefinita, gli utenti non amministrativi non dispongono delle autorizzazioni per creare database o tabelle all'interno di un database. La creazione del database è controllata a livello di account utilizzando le impostazioni di Lake Formation in modo che solo i responsabili autorizzati possano creare database. Per ulteriori informazioni, consulta Creazione di un database. Per creare una tabella, un principale richiede l'`CREATE_TABLE`autorizzazione sul database in cui viene creata la tabella. Per ulteriori informazioni, consulta Creazione di tabelle.

Autorizzazioni implicite ed esplicite

Lake Formation fornisce autorizzazioni implicite a seconda della persona e delle azioni che la persona compie. Ad esempio, gli amministratori del data lake ottengono automaticamente le `DESCRIBE` autorizzazioni per tutte le risorse all'interno del Data Catalog, le autorizzazioni per la localizzazione dei dati per tutte le posizioni, le autorizzazioni per creare database e tabelle in tutte le posizioni e le autorizzazioni per qualsiasi risorsa. `Grant` `Revoke` I creatori di database ottengono automaticamente tutte le autorizzazioni di database sui database che creano e i creatori di tabelle ottengono tutte le autorizzazioni sulle tabelle che creano. Per ulteriori informazioni, consulta Autorizzazioni implicite di Lake Formation.

Autorizzazioni concedibili

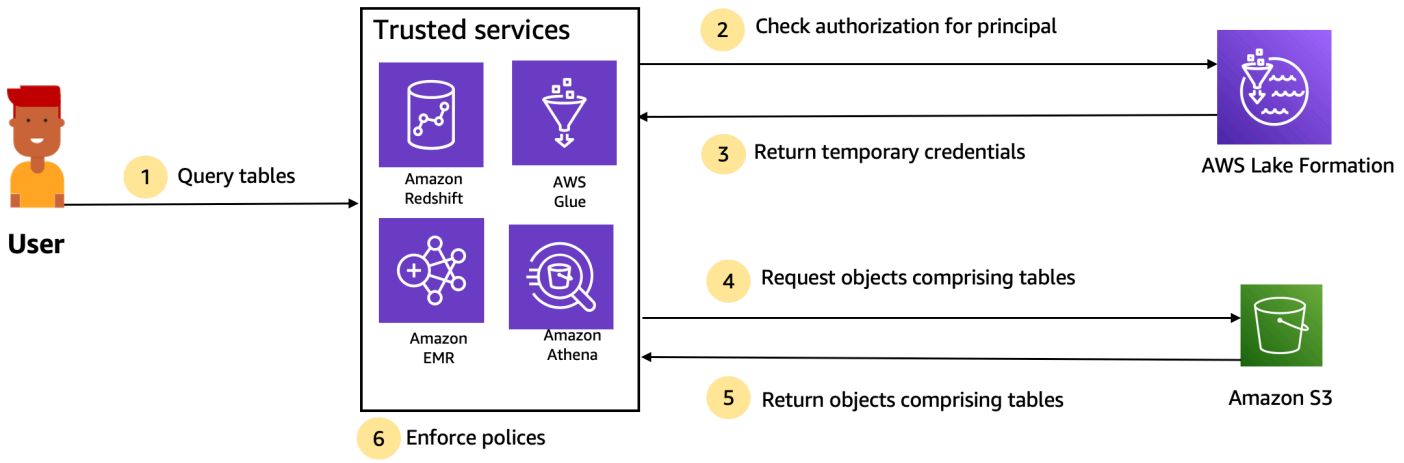
Gli amministratori di Data Lake hanno la possibilità di delegare la gestione delle autorizzazioni a utenti non amministrativi fornendo autorizzazioni concedibili. Quando a un committente vengono concesse autorizzazioni su una risorsa e un insieme di autorizzazioni, tale principale ottiene la capacità di concedere le autorizzazioni ad altri responsabili su quella risorsa.

Gestione degli accessi allo storage

Lake Formation utilizza la funzionalità di vendita di credenziali per fornire un accesso temporaneo ai dati di Amazon S3. La vendita di credenziali, o vendita di token, è uno schema comune che fornisce credenziali temporanee a utenti, servizi o altre entità allo scopo di concedere l'accesso a breve termine a una risorsa.

Lake Formation sfrutta questo modello per fornire un accesso a breve termine a servizi di AWS analisi come Athena per accedere ai dati per conto del committente chiamante. Quando concedono le autorizzazioni, gli utenti non devono aggiornare le policy dei bucket Amazon S3 o le policy IAM e non hanno bisogno dell'accesso diretto ad Amazon S3.

Il diagramma seguente mostra come Lake Formation fornisce l'accesso temporaneo alle località registrate:



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

1. Un principale (utente) inserisce una query o una richiesta di dati per una tabella tramite un servizio integrato affidabile come Athena, Amazon EMR, Redshift Spectrum o. AWS Glue
2. Il servizio integrato verifica l'autorizzazione di Lake Formation per la tabella e le colonne richieste e determina l'autorizzazione. Se l'utente non è autorizzato, Lake Formation nega l'accesso ai dati e la query ha esito negativo.
3. Una volta completata l'autorizzazione e attivata l'autorizzazione all'archiviazione per la tabella e l'utente, il servizio integrato recupera le credenziali temporanee da Lake Formation per accedere ai dati.
4. Il servizio integrato utilizza le credenziali temporanee di Lake Formation per richiedere oggetti da Amazon S3.
5. Amazon S3 fornisce gli oggetti Amazon S3 al servizio integrato. Gli oggetti Amazon S3 contengono tutti i dati della tabella.
6. Il servizio integrato esegue la necessaria applicazione delle politiche di Lake Formation, come il filtraggio a livello di colonna, a livello di riga e/o a livello di cella. Il servizio integrato elabora le interrogazioni e restituisce i risultati all'utente.

Abilita l'applicazione delle autorizzazioni a livello di storage per le tabelle del Data Catalog

Per impostazione predefinita, l'applicazione a livello di storage non è abilitata per le tabelle all'interno del Data Catalog. Per abilitare l'applicazione a livello di storage, devi registrare la posizione Amazon S3 dei tuoi dati di origine con Lake Formation e fornire un ruolo IAM. Le autorizzazioni a livello di storage verranno abilitate per tutte le tabelle con lo stesso percorso di posizione della tabella o lo stesso prefisso della posizione Amazon S3.

Quando un servizio integrato richiede l'accesso alla posizione dei dati per conto di un utente, il servizio Lake Formation assume questo ruolo e restituisce le credenziali al servizio richiesto con autorizzazioni limitate alla risorsa in modo che sia possibile effettuare l'accesso ai dati. Il ruolo IAM registrato deve disporre di tutti gli accessi necessari alla posizione Amazon S3, comprese AWS KMS le chiavi.

Per ulteriori informazioni, consulta Registrazione di una sede Amazon S3.

Servizi supportati AWS

AWSservizi di analisi come Athena, Redshift Spectrum, Amazon EMR Amazon QuickSight e integrazione Amazon SageMaker con Lake AWS Formation utilizzando le AWS Glue operazioni API di vendita delle credenziali di Lake Formation. Per un elenco completo dei AWS servizi che si integrano con Lake Formation e il livello di granularità e i formati di tabella che supportano, consulta. Collaborazione con altri AWS servizi

Condivisione dei dati tra account in Lake Formation

Con Lake Formation, puoi condividere le risorse del Data Catalog (database e tabelle) all'interno di un AWS account e tra account con una semplice configurazione utilizzando il metodo della risorsa denominata o i tag LF. Puoi condividere un intero database o selezionare tabelle da un database con qualsiasi principale IAM (ruoli e utenti IAM) in un account, con altri account a livello di AWS account o direttamente con i principali IAM in un altro account.

Puoi anche condividere le tabelle del Data Catalog con filtri di dati per limitare l'accesso ai dettagli a livello di riga e di cella. Lake Formation utilizza AWS Resource Access Manager (AWS RAM) per facilitare la concessione di autorizzazioni tra account. Quando una risorsa viene condivisa tra due account, AWS RAM invia gli inviti all'account del destinatario. Quando un utente accetta un invito alla AWS RAM condivisione, AWS RAM fornisce le autorizzazioni necessarie a Lake Formation per avere a disposizione le risorse del Data Catalog e abilitare l'applicazione a livello di storage. Per ulteriori informazioni, consulta Condivisione dei dati tra account in Lake Formation.

Quando l'amministratore del data lake dell'account del destinatario accetta la AWS RAM condivisione, le risorse condivise sono disponibili nell'account del destinatario. L'amministratore del data lake concede ulteriori autorizzazioni Lake Formation sulla risorsa condivisa ai principali IAM aggiuntivi nell'account del destinatario, se l'amministratore dispone delle `GRANTABLE` autorizzazioni sulla risorsa condivisa.

Tuttavia, i responsabili non possono interrogare le risorse condivise utilizzando Athena o Redshift Spectrum senza un collegamento alla risorsa. Un collegamento a una risorsa è un'entità nel Data Catalog ed è simile a un concetto Linux-Symlink.

L'amministratore del data lake dell'account del destinatario crea un link alla risorsa condivisa. L'amministratore concede `Describe` le autorizzazioni sul collegamento alla risorsa con le autorizzazioni richieste sulla risorsa condivisa originale ad altri utenti. Un utente nell'account destinatario può quindi utilizzare il collegamento alla risorsa per interrogare la risorsa condivisa utilizzando Athena e Redshift Spectrum. Per ulteriori informazioni sui collegamenti alle risorse, consulta. Creazione di collegamenti alle risorse

componenti di Lake Formation mation

AWS Lake Formationsi basa sull'interazione di diversi componenti per creare e gestire il tuo data lake.

Console Lake Formation mation

Utilizzi la console Lake Formation per definire e gestire il tuo data lake e concedere e revocare le autorizzazioni di Lake Formation. Puoi utilizzare i blueprint sulla console per scoprire, pulire, trasformare e inserire dati. Puoi anche abilitare o disabilitare l'accesso alla console per i singoli utenti di Lake Formation.

API Lake Formation e interfaccia a riga di comando

Lake Formation fornisce operazioni API tramite diversi SDK specifici del linguaggio e ilAWS Command Line Interface (AWS CLI). L'API Lake Formation funziona insieme all'AWS GlueAPI. L'API Lake Formation si concentra principalmente sulla gestione delle autorizzazioni di Lake Formation, mentre l'AWS GlueAPI fornisce un'API di catalogo dati e un'infrastruttura gestita per la definizione, la pianificazione e l'esecuzione delle operazioni ETL sui dati.

Per informazioni sull'AWS GlueAPI, consulta la Guida per gliAWS Glue sviluppatori. Per informazioni sull'utilizzo diAWS CLI, consulta consulta consulta consulta consulta consulta AWS CLILake Formation.

Altri servizi AWS

Lake Formation utilizza i seguenti servizi:

- AWS Glueper orchestrare job e crawler per trasformare i dati utilizzando leAWS Glue trasformazioni.
- IAM concederà politiche di autorizzazione ai dirigenti di Lake Formation. Il modello di autorizzazione Lake Formation amplia il modello di autorizzazione IAM per proteggere il tuo data lake.

Terminologia dei Lake Formation

Di seguito sono riportati alcuni termini importanti che incontrerai in questa guida.

Lago di dati

Il data lake è costituito dai tuoi dati persistenti archiviati in Amazon S3 e gestiti da Lake Formation utilizzando un Data Catalog. Un data lake in genere archivia quanto segue:

- Dati strutturati e non strutturati
- Dati grezzi e dati trasformati

Affinché un percorso Amazon S3 si trovi all'interno di un data lake, deve essere registrato presso Lake Formation.

Accesso ai dati

Lake Formation fornisce un accesso sicuro e granulare ai dati attraverso un nuovo modello di autorizzazioni di concessione/revoca che aumenta le politiche (IAM). AWS Identity and Access Management

Analisti e data scientist possono utilizzare l'intero portafoglio di servizi di AWS analisi e apprendimento automatico, come Amazon Athena, per accedere ai dati. Le politiche di sicurezza configurate di Lake Formation aiutano a garantire che gli utenti possano accedere solo ai dati a cui sono autorizzati ad accedere.

Modalità di accesso ibrida

La modalità di accesso ibrido consente di proteggere e accedere ai dati catalogati utilizzando sia le autorizzazioni Lake Formation che le autorizzazioni IAM e Amazon S3. La modalità di accesso ibrido consente agli amministratori dei dati di integrare le autorizzazioni di Lake Formation in modo selettivo e incrementale, concentrandosi su un caso d'uso del data lake alla volta.

Piano

Un blueprint è un modello di gestione dei dati che consente di inserire facilmente i dati in un data lake. Lake Formation fornisce diversi modelli, ciascuno per un tipo di sorgente predefinito, come un database relazionale o dei log. AWS CloudTrail Da un blueprint, puoi creare un flusso di lavoro. I flussi di lavoro sono costituiti da AWS Glue crawler, job e trigger generati per orchestrare il caricamento e l'aggiornamento dei dati. I blueprint utilizzano l'origine dei dati, la destinazione dei dati e la pianificazione come input per configurare il flusso di lavoro.

Flusso di lavoro

Un workflow è un contenitore per una serie di AWS Glue job, crawler e trigger correlati. Il flusso di lavoro viene creato in Lake Formation e questo viene eseguito nel AWS Glue servizio. Lake Formation può tracciare lo stato di un flusso di lavoro come singola entità.

Quando si definisce un flusso di lavoro, si seleziona il progetto su cui si basa. È quindi possibile eseguire flussi di lavoro su richiesta o in base a una pianificazione.

I flussi di lavoro creati in Lake Formation sono visibili nella AWS Glue console come grafo aciclico diretto (DAG). Utilizzando il DAG, è possibile tenere traccia dell'avanzamento del flusso di lavoro ed eseguire la risoluzione dei problemi.

Catalogo dati

Il Data Catalog è il tuo archivio di metadati persistente. Si tratta di un servizio gestito che ti consente di archiviare, annotare e condividere i metadati nel cloud AWS nello stesso modo in cui lo faresti nel metastore Apache Hive. Fornisce un repository uniforme in cui diversi sistemi possono archiviare e trovare metadati per tenere traccia dei dati nei silos di dati e quindi utilizzare tali metadati per interrogare e trasformare i dati. Lake Formation utilizza il AWS Glue Data Catalog per archiviare metadati su data lake, sorgenti di dati, trasformazioni e destinazioni.

I metadati sulle fonti di dati e sulle destinazioni sono sotto forma di database e tabelle. Le tabelle memorizzano informazioni sullo schema, informazioni sulla posizione e altro ancora. I database sono raccolte di tabelle. Lake Formation fornisce una gerarchia di autorizzazioni per controllare l'accesso a database e tabelle nel Data Catalog.

Ogni AWS account dispone di un catalogo dati per regione. AWS

Dati sottostanti

I dati sottostanti si riferiscono ai dati di origine o ai dati all'interno dei data lake a cui fanno riferimento le tabelle del Data Catalog.

Principale

Un principale è un utente o un ruolo AWS Identity and Access Management (IAM) o un utente di Active Directory.

Amministratore del data lake

Un amministratore del data lake è un responsabile che può concedere a qualsiasi principale (incluso se stesso) qualsiasi autorizzazione su qualsiasi risorsa o posizione dei dati del Data Catalog. Designare un amministratore del data lake come primo utente del Data Catalog. Questo utente può quindi concedere autorizzazioni più granulari sulle risorse ad altri responsabili.

> **Note**
>
> Gli utenti amministrativi IAM, ovvero gli utenti con la policy `AdministratorAccess` AWS gestita, non sono automaticamente amministratori di data lake. Ad esempio, non possono concedere le autorizzazioni di Lake Formation sugli oggetti del catalogo a meno che non abbiano ottenuto le autorizzazioni per farlo. Tuttavia, possono utilizzare la console o l'API di Lake Formation per designarsi amministratori di data lake.

Per informazioni sulle funzionalità di un amministratore di data lake, consulta. Autorizzazioni implicite di Lake Formation Per informazioni sulla designazione di un utente come amministratore del data lake, consultaCrea un amministratore del data lake.

AWSintegrazioni di servizi con Lake Formation

Puoi utilizzare Lake Formation per gestire le autorizzazioni di accesso a livello di database, tabelle e colonne sui dati archiviati in Amazon S3. Dopo aver registrato i dati con Lake Formation, puoi utilizzare servizi di AWS analisi come Amazon AthenaAWS Glue, Amazon Redshift Spectrum, Amazon EMR per interrogare i dati. I seguenti AWS servizi si integrano AWS Lake Formation e rispettano le autorizzazioni di Lake Formation.

Servizio AWS	Dettagli sull'integrazione
AWS Glue	Argomento di riferimento: AWS Lake Formation Utilizzo con AWS Glue AWS Gluee Lake Formation condividono lo stesso Data Catalog. Per le operazioni della console (come la visualizzazione di un elenco di tabelle) e tutte le operazioni API, AWS Glue gli utenti possono accedere solo ai database e alle tabelle su cui dispongono delle autorizzazioni Lake Formation.
Amazon Athena	Argomento di riferimento: Utilizzo AWS Lake Formation con Amazon Athena Usa Lake Formation per consentire o negare le autorizzazioni alla lettura dei dati in Amazon S3. Quando Amazon Athena gli utenti selezionano il AWS Glue catalogo nell'editor di query, possono interrogare solo i database, le tabelle e le colonne su cui dispongon o delle autorizzazioni Lake Formation. Le query che utilizzano i manifesti non sono supportate. Attualmente, Lake Formation non supporta la gestione delle autorizza zioni su operazioni di scrittura come `VACUUM` `UPDATE` e `OPTIMIZE` sulle tabelle in Open Table Formats. `MERGE` Oltre ai principali che si autenticano con Athena tramite AWS Identity and Access Management (IAM), Lake Formation supporta gli utenti Athena che si connettono tramite il driver JDBC o ODBC e si autenticano tramite SAML. I provider SAML supportati includono Okta e Microsoft Active Directory Federation Service (ADFS).
Amazon Redshift Spectrum	Argomento di riferimento: Utilizzo AWS Lake Formation con Amazon Redshift Spectrum Quando gli utenti di Amazon Redshift creano uno schema esterno su un database inAWS Glue Data Catalog, possono interrogare solo le tabelle e le colonne dello schema su cui dispongono delle autorizza zioni Lake Formation.

Servizio AWS	Dettagli sull'integrazione
Edizione Amazon QuickSight Enterprise	Riferimento: Utilizzo AWS Lake Formation con Amazon QuickSight Quando un utente di Amazon QuickSight Enterprise Edition esegue una query su un set di dati in una posizione Amazon S3, deve disporre dell'autorizzazione Lake SELECT Formation sui dati.
Amazon EMR	Riferimento: Utilizzo AWS Lake Formation con Amazon EMR Puoi integrare le autorizzazioni di Lake Formation quando crei un cluster Amazon EMR con un ruolo di runtime. Un ruolo di runtime è un ruolo IAM che associ ai job o alle query di Amazon EMR, quindi Amazon EMR utilizza questo ruolo per accedere alle risorse. AWS

Lake Formation collabora anche con AWS Key Management Service(AWS KMS) per consentirti di configurare più facilmente questi servizi integrati per crittografare e decrittografare i dati nelle sedi Amazon Simple Storage Service (Amazon S3).

Risorse aggiuntive per Lake Formation

Argomenti

- Blog
- Seminari e webinar tecnici
- Architettura moderna
- Risorse Data Mesh
- Guide di Best practice

Blog

- AWS Lake FormationRiepilogo dell'anno 2022
- Architettura di dati moderna multiregionale altamente resiliente
- Condivisione tra account tramite tag LF per indirizzare i responsabili IAM

- Dashboard dell'inventario delle autorizzazioni di Lake Formation
- Mesh di dati basato sugli eventi

Seminari e webinar tecnici

- re:Invent 2020 — Data lake: crea, proteggi e condividi facilmente conAWS Lake Formation
- re:Invent 2022 — Creazione e gestione di un datalake su Amazon S3
- AWSSummit SF 2022 — Comprendere e realizzare un'architettura dati moderna
- AWSSummit ATL 2022 — Data lake moderni conAWS Lake Formation Amazon Redshift eAWS Glue
- AWSSummit ANZ 2022 — Data lake, case sul lago e data mesh: cosa, perché e come?
- AWSColloqui tecnici online: semplificazione delle autorizzazioni e della governance nel tuo data lake

Architettura moderna

- Modelli architettonici moderni

Risorse Data Mesh

- Crea un'architettura di dati moderna e un modello di data mesh su larga scala utilizzando il controllo degli accessiAWS Lake Formation basato su tag
- In che modo JPMorgan Chase ha creato un'architettura data mesh per generare un valore significativo e migliorare la propria piattaforma dati aziendale
- Costruisci una rete di dati suAWS

Guide di Best practice

- AWS Lake Formationguide sulle migliori pratiche

Guida introduttiva a Lake Formation

Ti consigliamo di iniziare con le sezioni seguenti:

- AWS Lake Formation: Come funziona— Scopri la terminologia essenziale e il modo in cui i vari componenti interagiscono.
- Guida introduttiva a Lake Formation— Ottieni informazioni sui prerequisiti e completa importanti attività di configurazione.
- Tutorial— Segui step-by-step i tutorial per imparare a usare Lake Formation.
- Sicurezza in AWS Lake Formation— Scopri come puoi contribuire a proteggere l'accesso ai dati in Lake Formation.

Guida introduttiva a Lake Formation

Se non hai effettuato la registrazione a AWS o hai bisogno di aiuto per iniziare, completa le seguenti attività.

Argomenti

Completa i processi di configurazione iniziali AWS

Per utilizzare AWS Lake Formation, devi innanzitutto completare i seguenti passaggi:

Argomenti

Registrarsi per creare un Account AWS

Se non disponi di un Account AWS, completa la procedura seguente per crearne uno.

Per registrarsi a un Account AWS

1. Apri la pagina all'indirizzo https://portal.aws.amazon.com/billing/signup.
2. Segui le istruzioni online.

   Nel corso della procedura di registrazione riceverai una telefonata, durante la quale sarà necessario inserire un codice di verifica attraverso la tastiera del telefono.

   Durante la registrazione di un Account AWS, viene creato un Utente root dell'account AWS. L'utente root dispone dell'accesso a tutte le risorse e tutti i Servizi AWSnell'account. Come best practice di sicurezza, assegna l'accesso amministrativo a un utente amministrativo e utilizza solo l'utente root per eseguire attività che richiedono l'accesso di un utente root.

Al termine del processo di registrazione, riceverai un'e-mail di conferma da AWS. È possibile visualizzare l'attività corrente dell'account e gestire l'account in qualsiasi momento accedendo all'indirizzo https://aws.amazon.com/ e selezionando Il mio account.

Creazione di un utente amministratore

Dopo aver effettuato la registrazione di un Account AWS, proteggi Utente root dell'account AWS, abilita AWS IAM Identity Center e crea un utente amministratore in modo da non utilizzare l'utente root per le attività quotidiane.

Protezione dell'Utente root dell'account AWS

1. Accedi alla AWS Management Console come proprietario dell'account scegliendo Utente root e immettendo l'indirizzo email del Account AWS. Nella pagina successiva, inserisci la password.

   Per informazioni sull'accesso utilizzando un utente root, consulta la pagina Accesso come utente root della Guida per l'utente di Accedi ad AWS.
2. Abilita l'autenticazione a più fattori (MFA) per l'utente root.

   Per ricevere istruzioni, consulta Abilitazione di un dispositivo MFA virtuale per l'utente root dell'Account AWS (console) nella Guida per l'utente IAM.

Creazione di un utente amministratore

1. Abilita Centro identità IAM.

   Per istruzioni, consulta Abilitazione di AWS IAM Identity Center nella Guida per l'utente di AWS IAM Identity Center.
2. In Centro identità IAM, assegna l'accesso amministrativo a un utente amministratore.

   Per un tutorial sull'utilizzo di IAM Identity Center directory come origine di identità, consulta Configure user access with the default IAM Identity Center directory nella Guida per l'utente di AWS IAM Identity Center.

Accesso come utente amministratore

- Per accedere con l'utente IAM Identity Center, utilizza l'URL di accesso che è stato inviato al tuo indirizzo e-mail quando hai creato l'utente IAM Identity Center.

Per informazioni sull'accesso utilizzando un utente IAM Identity Center, consulta Accedere al portale di accesso AWS nella Guida per l'utente Accedi ad AWS.

Concessione dell'accesso programmatico

Gli utenti hanno bisogno di un accesso programmatico se desiderano interagire con AWS esternamente a AWS Management Console. La modalità con cui concedere l'accesso programmatico dipende dal tipo di utente che accede ad AWS.

Per fornire agli utenti l'accesso programmatico, scegli una delle seguenti opzioni.

Quale utente necessita dell'accesso programmatico?	Per	Come
Identità della forza lavoro (Utenti gestiti nel centro identità IAM)	Utilizza credenziali temporane e per firmare richieste programmatiche alla AWS CLI, agli SDK AWS o alle API AWS.	Segui le istruzioni per l'interfa ccia che desideri utilizzare. • Per la AWS CLI, consulta la pagina Configurazione della AWS CLI per l'uso di AWS IAM Identity Center nella Guida per l'utente dell'AWS Command Line Interface. • Per gli SDK AWS, gli strumenti e le API AWS, consulta la pagina Autentica zione Centro identità IAM nella Guida di riferimento per SDK e strumenti AWS.
IAM	Utilizza credenziali temporane e per firmare richieste programmatiche alla AWS CLI, agli SDK AWS o alle API AWS.	Segui le istruzioni in Utilizzo di credenziali temporanee con le risorse AWS nella Guida per l'utente IAM.

Quale utente necessita dell'accesso programmatico?	Per	Come
IAM	(Non consigliato) Utilizza credenziali a lungo termine per firmare richieste programmatiche alla AWS CLI, agli SDK AWS o alle API AWS.	Segui le istruzioni per l'interfaccia che desideri utilizzare. • Per la AWS CLI, consulta la pagina Autenticazione tramite credenziali utente IAM nella Guida per l'utente dell'AWS Command Line Interface. • Per gli SDK e gli strumenti AWS, consulta la pagina Autenticazione con credenziali a lungo termine nella Guida di riferimento per SDK e strumenti AWS. • Per le API AWS, consulta la pagina Gestione delle chiavi di accesso per utenti IAM nella Guida per l'utente IAM.

Configurazione di AWS Lake Formation

Le sezioni seguenti forniscono informazioni sulla configurazione di Lake Formation per la prima volta. Non tutti gli argomenti di questa sezione sono necessari per iniziare a usare Lake Formation. Puoi utilizzare le istruzioni per configurare il modello di autorizzazioni Lake Formation per gestire AWS Glue Data Catalog gli oggetti e le posizioni dei dati esistenti in Amazon Simple Storage Service (Amazon S3) Simple Storage Service (Amazon S3).

1. Crea un amministratore del data lake
2. Modifica il modello di autorizzazione predefinito o utilizza la modalità di accesso ibrida
3. the section called “Configura una posizione Amazon S3 per il tuo data lake”
4. the section called “Assegna le autorizzazioni agli utenti di Lake Formation”

5. the section called “Integrazione di IAM Identity Center”
6. the section called “(Facoltativo) Impostazioni di filtraggio dei dati esterni”
7. the section called “(Facoltativo) Concedi l'accesso alla chiave di crittografia Data Catalog”
8. (Facoltativo) Crea un ruolo IAM per i flussi di lavoro

Questa sezione mostra come configurare le risorse di Lake Formation in due modi diversi:

- Utilizzo del modello AWS CloudFormation
- Utilizzo della console Lake Formation

Per configurare Lake Formation tramite AWS console, vai aCrea un amministratore del data lake.

Configura le risorse di Lake Formation utilizzando il AWS CloudFormation modello

> **Note**
>
> Lo AWS CloudFormation stack esegue i passaggi da 1 a 6 di cui sopra, ad eccezione dei passaggi 2 e 5. Esegui Modifica il modello di autorizzazione predefinito o utilizza la modalità di accesso ibrida e the section called “Integrazione di IAM Identity Center” manualmente dalla console Lake Formation.

1. Accedi alla AWS CloudFormation console all'indirizzo https://console.aws.amazon.com/cloudformation come amministratore IAM nella regione Stati Uniti orientali (Virginia settentrionale).
2. Scegli Launch Stack.
3. Scegli Avanti nella schermata Crea stack.
4. Inserisci un nome per lo stack.
5. Per DatalakeAdminNamee DatalakeAdminPassword, inserisci il nome utente e la password per l'utente amministratore di Data Lake.
6. Per DatalakeUser1Name e DatalakeUser1Password, inserisci il nome utente e la password per l'utente analista di data lake.
7. Per DataLakeBucketName, inserisci il nuovo nome del bucket che verrà creato.

8. Seleziona Avanti.
9. Nella pagina successiva, scegli Avanti.
10. Controlla i dettagli nella pagina finale e seleziona Riconosco che AWS CloudFormation potrebbe creare risorse IAM.
11. Scegli Crea.

    La creazione dello stack può richiedere fino a due minuti.

Pulizia delle risorse

Se desideri ripulire le risorse dello AWS CloudFormation stack:

1. Annulla la registrazione del bucket Amazon S3 creato dallo stack e registrato come posizione di data lake.
2. Elimina lo stack. AWS CloudFormation Questo eliminerà tutte le risorse create dallo stack.

Crea un amministratore del data lake

Gli amministratori di Data Lake sono inizialmente gli unici utenti o ruoli AWS Identity and Access Management (IAM) che possono concedere le autorizzazioni di Lake Formation sulle posizioni dei dati e sulle risorse del Data Catalog a qualsiasi principale (incluso se stesso). Per ulteriori informazioni sulle funzionalità di amministratore del data lake, consulta. Autorizzazioni implicite di Lake Formation Per impostazione predefinita, Lake Formation consente di creare fino a 30 amministratori di data lake.

Puoi creare un amministratore di data lake utilizzando la console Lake Formation o il `PutDataLakeSettings` funzionamento dell'API Lake Formation.

Le seguenti autorizzazioni sono necessarie per creare un amministratore del data lake. L'`Administrator`utente dispone implicitamente di queste autorizzazioni.

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

Se concedi la `AWSLakeFormationDataAdmin` policy a un utente, quell'utente non sarà in grado di creare altri utenti amministratori di Lake Formation.

Per creare un amministratore del data lake (console)

1. Se l'utente che deve essere amministratore del data lake non esiste ancora, utilizza la console IAM per crearlo. Altrimenti, scegli un utente esistente che sarà l'amministratore del data lake.

   > **Note**
   >
   > Ti consigliamo di non selezionare un utente amministrativo IAM (utente con la policy `AdministratorAccess` AWS gestita) come amministratore del data lake.

   Allega le seguenti politiche AWS gestite all'utente:

Policy	Obbligatorio?	Note
`AWSLakeFormationDataAdmin`	Obbligatorio	Autorizzazioni di base per gli amministratori del data lake. Questa policy AWS gestita contiene un divieto esplicito per il funzionamento dell'API Lake Formation, `PutDataLakeSetting` che impedisce agli utenti di creare nuovi amministratori di data lake.
`AWSGlueConsoleFullAccess` , `CloudWatchLogsReadOnlyAccess`	Facoltativo	Allega queste politiche se l'amministratore del data lake intende risolvere i problemi dei flussi di lavoro creati dai blueprint di Lake Formation. Queste politiche consentono all'amministratore del data lake di visualizzare le informazioni sulla risoluzione dei problemi nella AWS Glue console e nella Amazon CloudWatch Logs console. Per informazioni sui flussi di lavoro, consultathe section called “Importazione di dati tramite flusso di lavoro”.

Policy	Obbligatorio?	Note
`AWSLakeFormationCrossAccountManager`	Facoltativo	Allega questa policy per consentire all'amministratore del data lake di concedere e revocare le autorizzazioni tra più account sulle risorse di Data Catalog. Per ulteriori informazioni, consulta Condivisione dei dati tra account in Lake Formation.
`AmazonAthenaFullAccess`	Facoltativo	Allega questa policy se l'amministratore del data lake eseguirà le query in. Amazon Athena

2. Allega la seguente policy in linea, che concede all'amministratore del data lake l'autorizzazione a creare il ruolo collegato al servizio Lake Formation. Un nome suggerito per la policy è. `LakeFormationSLR`

   Il ruolo collegato ai servizi consente all'amministratore del data lake di registrare più facilmente le sedi Amazon S3 con Lake Formation. Per ulteriori informazioni sul ruolo collegato ai servizi di Lake Formation, vedere. the section called “Uso di ruoli collegati ai servizi”

   > **Important**
   >
   > In tutti i criteri seguenti, sostituiscili <account-id>con un numero di AWS account valido.

```
{
    "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. (Facoltativo) Allega la seguente politica `PassRole` in linea all'utente. Questa policy consente all'amministratore del data lake di creare ed eseguire flussi di lavoro. L'`iam:PassRole`autorizzazione consente al flusso di lavoro di assumere il ruolo di `LakeFormationWorkflowRole` creare crawler e job e di assegnare il ruolo ai crawler e ai job creati. Un nome suggerito per la policy è. `UserPassRole`

> **Important**
>
> <account-id>Sostituiscilo con un numero di AWS conto valido.

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

4. (Facoltativo) Allega questa politica in linea aggiuntiva se il tuo account concederà o riceverà autorizzazioni Lake Formation per più account. Questa policy consente all'amministratore del data lake di visualizzare e accettare AWS Resource Access Manager (AWS RAM) gli inviti

alla condivisione di risorse. Inoltre, per gli amministratori del data lake dell'account di AWS Organizations gestione, la policy include l'autorizzazione ad abilitare le concessioni tra account alle organizzazioni. Per ulteriori informazioni, consulta Condivisione dei dati tra account in Lake Formation.

Un nome suggerito per la policy è. `RAMAccess`

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

5. Apri la AWS Lake Formation console all'indirizzo https://console.aws.amazon.com/lakeformation/ e accedi come utente amministratore che hai creato Creazione di un utente amministratore o come `AdministratorAccess` utente con policy AWS gestite dagli utenti.
6. Se viene visualizzata la finestra Welcome to Lake Formation, scegli l'utente IAM che hai creato o selezionato nello Step 1, quindi scegli Inizia.
7. Se non vedi la finestra Welcome to Lake Formation, esegui i seguenti passaggi per configurare un Lake Formation Administrator.

   a. Nel pannello di navigazione, in Amministratori, scegli Ruoli e attività amministrative. Nella sezione Amministratori di Data Lake della pagina della console, scegli Aggiungi.

   b. Nella finestra di dialogo Aggiungi amministratori, in Tipo di accesso, scegli Amministratore del data lake.

   c. Per gli utenti e i ruoli IAM, scegli l'utente IAM che hai creato o selezionato nel passaggio 1, quindi scegli Salva.

Modifica il modello di autorizzazione predefinito o utilizza la modalità di accesso ibrida

Lake Formation inizia con le impostazioni «Usa solo il controllo di accesso IAM» abilitate per la compatibilità con AWS Glue Data Catalog il comportamento esistente. Queste impostazioni ti consentono di gestire l'accesso ai tuoi dati nel data lake e ai relativi metadati tramite policy IAM e policy bucket di Amazon S3.

Per facilitare la transizione delle autorizzazioni del data lake da un modello IAM e Amazon S3 alle autorizzazioni Lake Formation, ti consigliamo di utilizzare la modalità di accesso ibrida per Data Catalog. Con la modalità di accesso ibrida, hai a disposizione un percorso incrementale in cui puoi abilitare le autorizzazioni di Lake Formation per un set specifico di utenti senza interrompere altri utenti o carichi di lavoro esistenti.

Per ulteriori informazioni, consulta Modalità di accesso ibrida.

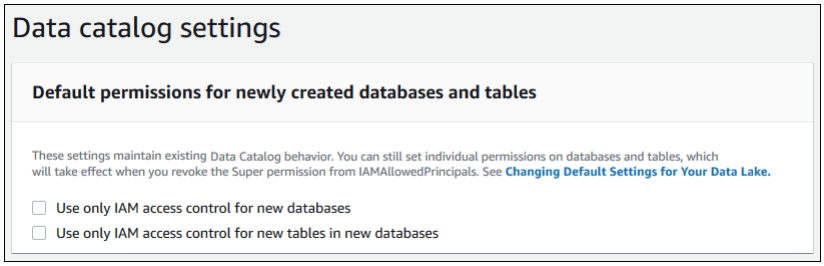
Disabilita le impostazioni predefinite per spostare tutti gli utenti esistenti di una tabella su Lake Formation in un unico passaggio.

> **Important**
>
> Se disponi di AWS Glue Data Catalog database e tabelle esistenti, non seguire le istruzioni in questa sezione. Seguire invece le istruzioni in the section called “Aggiornamento delle autorizzazioni per AWS Glue i dati al modello Lake Formation”.

> **Warning**
>
> Se disponi di un'automazione che crea database e tabelle nel Data Catalog, i passaggi seguenti potrebbero causare il fallimento dei processi di automazione e di estrazione, trasformazione e caricamento (ETL) a valle. Procedi solo dopo aver modificato i processi esistenti o concesso autorizzazioni esplicite di Lake Formation ai responsabili richiesti. Per informazioni sulle autorizzazioni di Lake Formation, vederethe section called “Riferimento alle autorizzazioni di Lake Formation”.

Per modificare le impostazioni predefinite del Data Catalog

1. Continua nella console di Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/. Assicurati di aver effettuato l'accesso come utente amministratore che hai creato Creazione di un utente amministratore o come utente con la policy `AdministratorAccess` AWS gestita.
2. Modifica le impostazioni del Data Catalog:
   a. Nel riquadro di navigazione, in Amministrazione, scegli Impostazioni Data Catalog.
   b. Deseleziona entrambe le caselle di controllo e scegli Salva.



3. Revoca l'`IAMAllowedPrincipals`autorizzazione per i creatori di database.
   a. Nel riquadro di navigazione, in Amministrazione, scegli Ruoli e attività amministrative.
   b. Nella pagina della console Ruoli e attività amministrative, nella sezione Creatori di database, seleziona il `IAMAllowedPrincipals` gruppo e scegli Revoca.

      Viene visualizzata la finestra di dialogo Revoca autorizzazioni, che indica che **`IAMAllowedPrincipals`** dispone dell'autorizzazione Crea database.
   c. Scegliete Revoke.

Assegna le autorizzazioni agli utenti di Lake Formation

Crea un utente per avere accesso al data lake in. AWS Lake Formation Questo utente dispone dei privilegi minimi per interrogare il data lake.

Per ulteriori informazioni sulla creazione di utenti o gruppi, consulta le identità IAM nella IAM User Guide.

Per assegnare a un utente non amministratore le autorizzazioni per accedere ai dati di Lake Formation

1. Apri la console IAM all'indirizzo https://console.aws.amazon.com/iam e accedi come utente amministratore che hai creato Creazione di un utente amministratore o come utente con la policy `AdministratorAccess` AWS gestita.
2. Scegli Utenti o Gruppi di utenti.
3. Nell'elenco, scegli il nome dell'utente o del gruppo in cui integrare una policy.

   Seleziona Autorizzazioni.
4. Scegli Aggiungi autorizzazioni e scegli Allega direttamente le politiche. Entra `Athena` nel campo di testo Filtra le politiche. Nell'elenco dei risultati, seleziona la casella per`AmazonAthenaFullAccess`.
5. Scegli il pulsante Crea politica. Nella pagina Crea policy, scegli la scheda JSON. Copia e incolla il seguente codice nell'editor delle politiche.

```
{
    "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. Scegli il pulsante Avanti in basso fino a visualizzare la pagina di revisione della politica. Inserisci un nome per la politica, ad esempio`DatalakeUserBasic`. Scegli Crea politica, quindi chiudi la scheda Politiche o la finestra del browser.

Configura una posizione Amazon S3 per il tuo data lake

Per utilizzare Lake Formation per gestire e proteggere i dati nel tuo data lake, devi prima registrare una sede Amazon S3. Quando registri una posizione, vengono registrati quel percorso Amazon S3 e tutte le cartelle in quel percorso, il che consente a Lake Formation di applicare le autorizzazioni a livello di storage. Quando l'utente richiede dati da un motore integrato come Amazon Athena, Lake Formation fornisce l'accesso ai dati anziché utilizzare le autorizzazioni degli utenti.

Quando registri una posizione, specifichi un ruolo IAM che concede autorizzazioni di lettura/scrittura su quella posizione. Lake Formation assume questo ruolo quando fornisce credenziali temporanee a AWS servizi integrati che richiedono l'accesso ai dati nella posizione registrata di Amazon S3. Puoi specificare il ruolo collegato ai servizi (SLR) di Lake Formation o creare il tuo ruolo.

Utilizzate un ruolo personalizzato nelle seguenti situazioni:

- Prevedi di pubblicare i parametri in Amazon CloudWatch Logs. Il ruolo definito dall'utente deve includere una politica per l'aggiunta di log in CloudWatch Logs e la pubblicazione di metriche oltre alle autorizzazioni SLR. Per un esempio di politica in linea che concede le autorizzazioni necessarie, vedi. CloudWatch Requisiti per i ruoli utilizzati per registrare le sedi
- La posizione Amazon S3 si trova in un account diverso. Per informazioni dettagliate, vedi the section called “Registrazione di una sede Amazon S3 in un altro account AWS”.
- La posizione Amazon S3 contiene dati crittografati con un. Chiave gestita da AWS Per informazioni dettagliate, consulta Registrazione di una posizione Amazon S3 crittografata e Registrazione di una posizione Amazon S3 crittografata tra più account AWS.
- Prevedi di accedere alla posizione Amazon S3 utilizzando Amazon EMR. Per ulteriori informazioni sui requisiti dei ruoli, consulta IAM roles for Lake Formation nella Amazon EMR Management Guide.

Il ruolo scelto deve disporre delle autorizzazioni necessarie, come descritto in. Requisiti per i ruoli utilizzati per registrare le sedi Per istruzioni su come registrare una sede Amazon S3, consulta. Aggiungere una posizione Amazon S3 al tuo data lake

(Facoltativo) Impostazioni di filtraggio dei dati esterni

Se intendi analizzare ed elaborare i dati nel tuo data lake utilizzando motori di query di terze parti, devi attivare l'opzione per consentire ai motori esterni di accedere ai dati gestiti da Lake Formation. Se non effettui l'attivazione, i motori esterni non saranno in grado di accedere ai dati nelle sedi Amazon S3 registrate con Lake Formation.

Lake Formation supporta le autorizzazioni a livello di colonna per limitare l'accesso a colonne specifiche di una tabella. Servizi di analisi integrati come Amazon Athena Amazon Redshift Spectrum e Amazon EMR recuperano i metadati delle tabelle non filtrati da. AWS Glue Data Catalog L'effettivo filtraggio delle colonne nelle risposte alle query è responsabilità del servizio integrato. È responsabilità degli amministratori di terze parti gestire correttamente le autorizzazioni per evitare l'accesso non autorizzato ai dati.

Attivare l'autorizzazione per consentire ai motori di terze parti di accedere e filtrare i dati (console)

1. Continua nella console di Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/. Assicurati di aver effettuato l'accesso come principale con l'autorizzazione IAM sull'operazione dell'`PutDataLakeSettings`API Lake Formation. L'utente amministratore IAM in cui hai creato Registrarsi per creare un Account AWS dispone di questa autorizzazione.
2. Nel riquadro di navigazione, in Amministrazione, scegli Impostazioni di integrazione delle applicazioni.
3. Nella pagina delle impostazioni di integrazione dell'applicazione, procedi come segue:
    a. Seleziona la casella Consenti ai motori esterni di filtrare i dati nelle località Amazon S3 registrate con Lake Formation.
    b. Inserisci i valori dei tag di sessione definiti per i motori di terze parti.
    c. Per gli ID AWS account, inserisci gli ID account da cui i motori di terze parti possono accedere alle località registrate con Lake Formation. Premi Invio dopo ogni ID account.
    d. Selezionare Salva.

Per consentire ai motori esterni di accedere ai dati senza la convalida dei tag di sessione, vedi Integrazione delle applicazioni per l'accesso completo alla tabella

(Facoltativo) Concedi l'accesso alla chiave di crittografia Data Catalog

Se AWS Glue Data Catalog è crittografato, concedi le autorizzazioni AWS Identity and Access Management (IAM) sulla AWS KMS chiave a tutti i responsabili che devono concedere le autorizzazioni di Lake Formation sui database e sulle tabelle del Data Catalog.

Per ulteriori informazioni, consulta la Guida per gli sviluppatori di AWS Key Management Service.

(Facoltativo) Crea un ruolo IAM per i flussi di lavoro

ConAWS Lake Formation, puoi importare i tuoi dati utilizzando flussi di lavoro eseguiti dai AWS Glue crawler. Un flusso di lavoro definisce l'origine dei dati e la pianificazione per l'importazione dei dati nel data lake. Puoi definire facilmente i flussi di lavoro utilizzando i blueprint o i modelli forniti da Lake Formation.

Quando crei un flusso di lavoro, devi assegnargli un ruolo AWS Identity and Access Management (IAM) che conceda a Lake Formation le autorizzazioni necessarie per importare i dati.

La procedura seguente presuppone la familiarità con IAM.

Per creare un ruolo IAM per i flussi di lavoro

1. Apri la console IAM all'indirizzo https://console.aws.amazon.com/iam e accedi come utente amministratore che hai creato Creazione di un utente amministratore o come utente con la policy `AdministratorAccess` AWS gestita.
2. Nel pannello di navigazione, scegli Ruoli, quindi Crea ruolo.
3. Nella pagina Crea ruolo, scegli AWSservice, quindi scegli Glue. Seleziona Avanti.
4. Nella pagina Aggiungi autorizzazioni, cerca la politica AWSGlueServiceRolegestita e seleziona la casella di controllo accanto al nome della politica nell'elenco. Quindi completa la procedura guidata di creazione del ruolo, assegnando un nome al ruolo. `LakeFormationWorkflowRole` Per finire, scegli Crea ruolo.
5. Torna alla pagina Ruoli, cerca `LakeFormationWorkflowRole` e scegli il nome del ruolo.
6. Nella pagina di riepilogo del ruolo, nella scheda Autorizzazioni, scegli Crea policy in linea. Nella schermata Crea policy, vai alla scheda JSON e aggiungi la seguente policy in linea. Un nome suggerito per la policy è. `LakeFormationWorkflow`

> **Important**
>
> Nella seguente politica, sostituiscila <account-id>con un Account AWS numero valido.

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

Di seguito sono riportate brevi descrizioni delle autorizzazioni contenute in questa politica:

- `lakeformation:GetDataAccess`consente ai lavori creati dal flusso di lavoro di scrivere nella posizione di destinazione.
- `lakeformation:GrantPermissions`consente al flusso di lavoro di concedere l'`SELECT`autorizzazione sulle tabelle di destinazione.
- `iam:PassRole`consente al servizio di assumere il ruolo di `LakeFormationWorkflowRole` creare crawler e job (istanze di flussi di lavoro) e di assegnare il ruolo ai crawler e ai job creati.

7. Verifica che al ruolo siano associate due policy. `LakeFormationWorkflowRole`
8. Se stai importando dati che si trovano all'esterno della posizione del data lake, aggiungi una policy in linea che conceda le autorizzazioni per leggere i dati di origine.

Aggiornamento delle autorizzazioni per AWS Glue i dati al modello AWS Lake Formation

AWS Lake Formationle autorizzazioni consentono un controllo granulare degli accessi per i dati nel data lake. Puoi utilizzare il modello di autorizzazioni Lake Formation per gestire AWS Glue Data Catalog gli oggetti e le posizioni dei dati esistenti in Amazon Simple Storage Service (Amazon S3).

Il modello di autorizzazioni Lake Formation utilizza autorizzazioni a grana grossa AWS Identity and Access Management (IAM) per l'accesso ai servizi API. Limita i dati a cui gli utenti e tali servizi possono accedere tramite la funzionalità Lake Formation. In confronto, il AWS Glue modello garantisce l'accesso ai dati tramite autorizzazioni IAM di controllo degli accessi granulari. Per effettuare il passaggio, segui i passaggi di questa guida.

Per ulteriori informazioni, consulta Panoramica delle autorizzazioni di Lake Formation .

Argomenti

Informazioni sull'aggiornamento al modello di autorizzazioni Lake Formation

Per mantenere la compatibilità con le versioni precedentiAWS Glue, per impostazione predefinita, AWS Lake Formation concede l'`Super`autorizzazione al `IAMAllowedPrincipals` gruppo su tutte le risorse del AWS Glue Data Catalog esistenti e concede l'`Super`autorizzazione sulle nuove risorse del Data Catalog se le impostazioni di controllo di accesso Use only IAM sono abilitate. Ciò fa sì che l'accesso alle risorse del Data Catalog e alle sedi Amazon S3 sia controllato esclusivamente da policy AWS Identity and Access Management (IAM). Il `IAMAllowedPrincipals` gruppo include tutti gli utenti e i ruoli IAM a cui è consentito l'accesso agli oggetti del Data Catalog dalle politiche IAM.

L'`Super`autorizzazione consente a un principale di eseguire tutte le operazioni di Lake Formation supportate sul database o sulla tabella su cui è concessa.

Puoi iniziare a utilizzare Lake Formation per gestire l'accesso ai tuoi dati registrando le posizioni delle risorse Data Catalog esistenti in Lake Formation o utilizzando la modalità di accesso ibrida. Quando registri una posizione Amazon S3 in modalità di accesso ibrido, puoi abilitare le autorizzazioni Lake Formation optando per i principali database e tabelle in quella posizione.

Per facilitare la transizione delle autorizzazioni del data lake da un modello IAM e Amazon S3 alle autorizzazioni Lake Formation, ti consigliamo di utilizzare la modalità di accesso ibrida per Data Catalog. Con la modalità di accesso ibrida, hai a disposizione un percorso incrementale in cui puoi abilitare le autorizzazioni di Lake Formation per un set specifico di utenti senza interrompere altri utenti o carichi di lavoro esistenti.

Per ulteriori informazioni, consulta Modalità di accesso ibrida.

Disabilita le impostazioni predefinite del Data Catalog per spostare tutti gli utenti esistenti di una tabella su Lake Formation in un unico passaggio.

Per iniziare a utilizzare le autorizzazioni di Lake Formation con i database e le tabelle AWS Glue Data Catalog esistenti, devi fare quanto segue:

1. Determina le autorizzazioni IAM esistenti degli utenti per ogni database e tabella.
2. Replica queste autorizzazioni in Lake Formation.
3. Per ogni posizione Amazon S3 che contiene dati:
   a. Revoca l'`Super`autorizzazione del `IAMAllowedPrincipals` gruppo su ogni risorsa del catalogo dati che fa riferimento a quella posizione.
   b. Registra la posizione con Lake Formation.
4. Pulisci le politiche IAM di controllo degli accessi a grana fine esistenti.

> **Important**
>
> Per aggiungere nuovi utenti durante il processo di transizione del Data Catalog, devi configurare le AWS Glue autorizzazioni granulari in IAM come in precedenza. È inoltre necessario replicare tali autorizzazioni in Lake Formation come descritto in questa sezione. Se i nuovi utenti dispongono delle politiche IAM dettagliate descritte in questa guida,

possono elencare tutti i database o le tabelle a cui è stata concessa l'autorizzazione. `Super IAMAllowedPrincipals` Possono anche visualizzare i metadati di tali risorse.

Segui i passaggi in questa sezione per eseguire l'aggiornamento al modello di autorizzazioni Lake Formation. Inizia consultando the section called “Passaggio 1: Elenca le autorizzazioni esistenti”.

Passaggio 1: Elenca le autorizzazioni esistenti di utenti e ruoli

Per iniziare a utilizzare AWS Lake Formation le autorizzazioni con i AWS Glue database e le tabelle esistenti, devi prima determinare le autorizzazioni esistenti degli utenti.

Important

Prima di iniziare, assicurati di aver completato le attività in. Nozioni di base

Argomenti

- Utilizzo dell'operazione API
- Utilizzo di AWS Management Console
- Uso di AWS CloudTrail

Utilizzo dell'operazione API

Utilizza l'operazione ListPoliciesGrantingServiceAccessAPI AWS Identity and Access Management (IAM) per determinare le policy IAM associate a ciascun principale (utente o ruolo). In base alle policy restituite nei risultati, puoi determinare le autorizzazioni IAM concesse al principale. È necessario richiamare l'API per ogni principale separatamente.

Example

L'AWS CLIesempio seguente restituisce le politiche allegate all'utente`glue_user1`.

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

Il comando restituisce risultati simili ai seguenti.

```
{
```

```
    "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
}
```

Utilizzo di AWS Management Console

Puoi anche visualizzare queste informazioni nella console AWS Identity and Access Management (IAM), nella scheda Access Advisor nella pagina di riepilogo dell'utente o del ruolo:

1. Apri la console IAM all'indirizzo https://console.aws.amazon.com/iam/.
2. Nel riquadro di navigazione, selezionare Users (Utenti) o Roles (Ruoli).
3. Scegli un nome nell'elenco per aprire la relativa pagina di riepilogo e scegli la scheda Access Advisor.
4. Ispeziona ciascuna politica per determinare la combinazione di database, tabelle e azioni per cui ogni utente dispone delle autorizzazioni.

   Ricorda di controllare i ruoli oltre agli utenti durante questo processo, perché i tuoi processi di elaborazione dei dati potrebbero assumere ruoli per accedere ai dati.

Uso di AWS CloudTrail

Un altro modo per determinare le autorizzazioni esistenti consiste nel AWS CloudTrail cercare le chiamate AWS Glue API in cui il `additionaleventdata` campo dei log contiene una voce.

`insufficientLakeFormationPermissions` Questa voce elenca il database e la tabella su cui l'utente necessita delle autorizzazioni di Lake Formation per eseguire la stessa azione.

Si tratta di registri di accesso ai dati, quindi non è garantito che producano un elenco completo degli utenti e delle relative autorizzazioni. Ti consigliamo di scegliere un ampio intervallo di tempo per registrare la maggior parte dei modelli di accesso ai dati degli utenti, ad esempio diverse settimane o mesi.

Per ulteriori informazioni, consulta Visualizzazione degli eventi con cronologia degli CloudTrail eventi nella Guida per l'AWS CloudTrailutente.

Successivamente, puoi impostare le autorizzazioni di Lake Formation in modo che corrispondano alle AWS Glue autorizzazioni. Per informazioni, consulta Fase 2: Impostare le autorizzazioni equivalenti di Lake Formation.

Fase 2: Impostare le autorizzazioni equivalenti di Lake Formation

Utilizzando le informazioni raccolte inPassaggio 1: Elenca le autorizzazioni esistenti di utenti e ruoli, concedi le AWS Lake Formation autorizzazioni corrispondenti alle autorizzazioni. AWS Glue Utilizzate uno dei seguenti metodi per eseguire le concessioni:

- Usa la console Lake Formation o ilAWS CLI.

  Per informazioni, consulta the section called “Concessione e revoca delle autorizzazioni di Data Catalog”.
- Usa le operazioni `GrantPermissions` o `BatchGrantPermissions` API.

  Per informazioni, consulta API per le autorizzazioni.

Per ulteriori informazioni, consulta Panoramica delle autorizzazioni di Lake Formation .

Dopo aver impostato le autorizzazioni di Lake Formation, procedi aPassaggio 3: concedere agli utenti le autorizzazioni IAM per utilizzare Lake Formation.

Passaggio 3: concedere agli utenti le autorizzazioni IAM per utilizzare Lake Formation

Per utilizzare il modello di AWS Lake Formation autorizzazioni, i responsabili devono disporre delle autorizzazioni AWS Identity and Access Management (IAM) sulle API Lake Formation.

Crea la seguente policy in IAM e collegala a ogni utente che ha bisogno di accedere al tuo data lake. Assegnare un nome alla policy `LakeFormationDataAccess`.

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

Successivamente, esegui l'upgrade alle autorizzazioni di Lake Formation una posizione dati alla volta. Per informazioni, consulta Passaggio 4: Passa i tuoi archivi dati al modello di autorizzazioni Lake Formation.

Passaggio 4: Passa i tuoi archivi dati al modello di autorizzazioni Lake Formation

Passa alle autorizzazioni di Lake Formation una posizione dati alla volta. A tale scopo, ripeti l'intera sezione fino a registrare tutti i percorsi di Amazon Simple Storage Service (Amazon S3) a cui fa riferimento il tuo Data Catalog.

Argomenti

- Verifica le autorizzazioni di Lake Formation
- Proteggi le risorse del Data Catalog esistenti
- Attiva le autorizzazioni di Lake Formation per la tua sede Amazon S3

Verifica le autorizzazioni di Lake Formation

Prima di registrare una sede, esegui una procedura di verifica per assicurarti che i mandanti corretti dispongano delle autorizzazioni Lake Formation richieste e che non vengano concesse autorizzazioni Lake Formation ai mandanti che non dovrebbero averle. Utilizzando l'operazione

`GetEffectivePermissionsForPath` API Lake Formation, identifica le risorse del Data Catalog che fanno riferimento alla posizione Amazon S3, insieme ai responsabili che dispongono delle autorizzazioni su tali risorse.

L'AWS CLIesempio seguente restituisce i database e le tabelle del catalogo dati che fanno riferimento al bucket Amazon S3. `products`

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

Nota l'opzione. `profile` Ti consigliamo di eseguire il comando come amministratore del data lake.

Di seguito è riportato un estratto dei risultati restituiti.

```
{
        "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 il tuo AWS Glue Data Catalog è crittografato, `GetEffectivePermissionsForPath` restituisce solo database e tabelle che sono stati creati o modificati dopo la disponibilità generale di Lake Formation.

Proteggi le risorse del Data Catalog esistenti

Successivamente, revoca l'`Super`autorizzazione da `IAMAllowedPrincipals` ogni tabella e database che hai identificato per la posizione.

> **Warning**
>
> Se disponi di un'automazione che crea database e tabelle nel Data Catalog, i passaggi seguenti potrebbero causare il fallimento dei processi di automazione e di estrazione, trasformazione e caricamento (ETL) a valle. Procedi solo dopo aver modificato i processi esistenti o concesso autorizzazioni esplicite di Lake Formation ai responsabili richiesti. Per informazioni sulle autorizzazioni di Lake Formation, vederethe section called “Riferimento alle autorizzazioni di Lake Formation”.

Per revocare un account **`Super`** da un tavolo **`IAMAllowedPrincipals`**

1. Apri la AWS Lake Formation console all'indirizzo https://console.aws.amazon.com/lakeformation/. Accedi come amministratore del data lake.
2. Nel pannello di navigazione, seleziona Tabelle.
3. Nella pagina Tabelle, seleziona il pulsante di opzione accanto alla tabella desiderata.
4. Nel menu Azioni, scegli Revoca.
5. Nella finestra di dialogo Revoca le autorizzazioni, nell'elenco Utenti e ruoli IAM, scorri verso il basso fino all'intestazione Gruppo e scegli IAM. AllowedPrincipals
6. Nella sezione Autorizzazioni della tabella, assicurati che Super sia selezionato, quindi scegli Revoke.

Per revocare da **`Super`** un database **`IAMAllowedPrincipals`**

1. Apri la AWS Lake Formation console all'indirizzo https://console.aws.amazon.com/lakeformation/. Accedi come amministratore del data lake.
2. Nel riquadro di navigazione, scegli Databases (Database).
3. Nella pagina Database, seleziona il pulsante di opzione accanto al database desiderato.
4. Nel menu Operazioni, scegliere Modifica.
5. Nella pagina Modifica database, deseleziona Usa solo il controllo di accesso IAM per le nuove tabelle in questo database, quindi scegli Salva.

6. Tornando alla pagina Database, assicurati che il database sia ancora selezionato, quindi nel menu Azioni, scegli Revoca.
7. Nella finestra di dialogo Revoca le autorizzazioni, nell'elenco Utenti e ruoli IAM, scorri verso il basso fino all'intestazione Gruppo e scegli IAM. AllowedPrincipals
8. In Autorizzazioni del database, assicurati che Super sia selezionato, quindi scegli Revoke.

Attiva le autorizzazioni di Lake Formation per la tua sede Amazon S3

Quindi, registra la sede Amazon S3 con Lake Formation. A tale scopo, puoi utilizzare la procedura descritta inAggiungere una posizione Amazon S3 al tuo data lake. In alternativa, utilizzate l'operazione `RegisterResource` API come descritto inAPI di vendita di credenziali.

> **Note**
>
> Se la sede di un genitore è registrata, non è necessario registrare le sedi dei figli.

Dopo aver completato questi passaggi e aver verificato che i tuoi utenti possano accedere ai propri dati, hai eseguito con successo l'upgrade alle autorizzazioni di Lake Formation. Continua con il passaggio successivo,. Fase 5: Proteggi le nuove risorse del Data Catalog

Fase 5: Proteggi le nuove risorse del Data Catalog

Successivamente, proteggi tutte le nuove risorse del Data Catalog modificando le impostazioni predefinite del Data Catalog. Disattiva le opzioni per utilizzare solo il controllo di accesso AWS Identity and Access Management (IAM) per nuovi database e tabelle.

> **Warning**
>
> Se disponi di un'automazione che crea database e tabelle nel Data Catalog, i passaggi seguenti potrebbero causare il fallimento dei processi di automazione e di estrazione, trasformazione e caricamento (ETL) a valle. Procedi solo dopo aver modificato i processi esistenti o concesso autorizzazioni esplicite di Lake Formation ai responsabili richiesti. Per informazioni sulle autorizzazioni di Lake Formation, vederethe section called “Riferimento alle autorizzazioni di Lake Formation”.

Per modificare le impostazioni predefinite del Data Catalog

1. Apri la AWS Lake Formation console all'indirizzo https://console.aws.amazon.com/lakeformation/. Accedi come utente amministrativo IAM (l'utente `Administrator` o un altro utente con la policy `AdministratorAccess` AWS gestita).
2. Nel pannello di navigazione scegli Settings (Impostazioni).
3. Nella pagina delle impostazioni del catalogo dati, deseleziona entrambe le caselle di controllo, quindi scegli Salva.

Il passaggio successivo consiste nel concedere agli utenti l'accesso a database o tabelle aggiuntivi in futuro. Per informazioni, consulta Fase 6: Offri agli utenti una nuova policy IAM per i futuri accessi ai data lake.

Fase 6: Offri agli utenti una nuova policy IAM per i futuri accessi ai data lake

Per concedere ai tuoi utenti l'accesso a database o tabelle di Data Catalog aggiuntivi in futuro, devi fornire loro la policy in linea a grana grossa AWS Identity and Access Management (IAM) che segue. Assegnare un nome alla policy `GlueFullReadAccess`.

> **Important**
>
> Se alleghi questa policy a un utente prima di revocarla `Super` `IAMAllowedPrincipals` su ogni database e tabella del tuo Data Catalog, quell'utente può visualizzare tutti i metadati di qualsiasi risorsa a cui è concesso. `Super` `IAMAllowedPrincipals`

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

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

> **Note**
>
> Le politiche in linea indicate in questo passaggio e nei passaggi precedenti contengono autorizzazioni IAM minime. Per le policy suggerite per gli amministratori di data lake, gli analisti di dati e altri personaggi, consulta. the section called “Riferimento ai personaggi di Lake Formation e alle autorizzazioni IAM”

Quindi, procedi a. Fase 7: Pulisci le policy IAM esistenti

Fase 7: Pulisci le policy IAM esistenti

Dopo aver configurato le AWS Lake Formation autorizzazioni e aver creato e collegato le politiche di controllo degli accessi AWS Identity and Access Management (IAM) granulari, completa il seguente passaggio finale:

- Rimuovi da utenti, gruppi e ruoli le vecchie policy IAM di controllo degli accessi granulari che hai replicato in Lake Formation.

In questo modo, ti assicuri che tali responsabili non abbiano più accesso diretto ai dati in Amazon Simple Storage Service (Amazon S3). Puoi quindi gestire l'accesso al data lake per questi principali interamente tramite Lake Formation.

AWS Lake Formation ed endpoint VPC dell'interfaccia (AWS PrivateLink)

Amazon VPC è un servizio AWS che puoi utilizzare per avviare risorse AWS in una rete virtuale da te definita. Con un VPC;, detieni il controllo delle impostazioni della rete, come l'intervallo di indirizzi IP, le sottoreti, le tabelle di routing e i gateway di rete.

Se utilizzi Amazon Virtual Private Cloud (Amazon VPC) per ospitare AWS le tue risorse, puoi stabilire una connessione privata tra il tuo VPC e Lake Formation. Utilizzi questa connessione in modo che Lake Formation possa comunicare con le risorse del tuo VPC senza passare attraverso la rete Internet pubblica.

Puoi stabilire una connessione privata tra VPC e AWS Lake Formation creando un endpoint VPC dell'interfaccia. Gli endpoint di interfaccia sono alimentati da AWS PrivateLink, una tecnologia che consente di accedere in modo privato alle API di Lake Formation senza un gateway Internet, un dispositivo NAT, una connessione VPN o una connessione. AWS Direct Connect Le istanze del tuo VPC non necessitano di indirizzi IP pubblici per comunicare con le API Lake Formation. Il traffico tra il tuo VPC e Lake Formation non esce dalla rete Amazon.

Ogni endpoint dell'interfaccia è rappresentato da una o più interfacce di rete elastiche nelle sottoreti.

Per ulteriori informazioni, consultare Endpoint VPC di interfaccia (AWS PrivateLink) nella Guida per l'utente di Amazon VPC.

Considerazioni sugli endpoint VPC di Lake Formation

Prima di configurare un endpoint VPC di interfaccia per Lake Formation, assicurati di esaminare le proprietà e le limitazioni degli endpoint dell'interfaccia nella Amazon VPC User Guide.

Lake Formation supporta l'esecuzione di chiamate a tutte le sue azioni API dal tuo VPC. Puoi utilizzare Lake Formation con endpoint VPC in tutti gli endpoint Regioni AWS che supportano sia gli endpoint Lake Formation che Amazon VPC.

Creazione di un endpoint VPC di interfaccia per Lake Formation

Puoi creare un endpoint VPC per il servizio Lake Formation utilizzando la console Amazon VPC o il (). AWS Command Line Interface AWS CLI Per ulteriori informazioni, consulta Creazione di un endpoint dell'interfaccia nella Guida per l'utente di Amazon VPC.

Crea un endpoint VPC per Lake Formation utilizzando il seguente nome di servizio:

- com.amazonaws.*region*.lakeformation

Se abiliti il DNS privato per l'endpoint, puoi effettuare richieste API a Lake Formation utilizzando il nome DNS predefinito per la regione, ad esempio. `lakeformation.us-east-1.amazonaws.com`

Per ulteriori informazioni, consulta Accesso a un servizio tramite un endpoint dell'interfaccia in Guida per l'utente di Amazon VPC.

Creazione di una policy sugli endpoint VPC per Lake Formation

Lake Formation supporta le policy degli endpoint VPC. Una policy degli endpoint VPC è una policy delle risorse AWS Identity and Access Management (IAM) che si collega a un endpoint quando si crea o si modifica l'endpoint.

Puoi allegare una policy per gli endpoint al tuo endpoint VPC che controlla l'accesso a Lake Formation. La policy specifica le informazioni riportate di seguito:

- Il principale che può eseguire operazioni.
- Le azioni che possono essere eseguite.
- Le risorse sui cui si possono eseguire operazioni.

Per ulteriori informazioni, consulta Controllo degli accessi ai servizi con endpoint VPC in Guida per l'utente di Amazon VPC.

Esempio: policy sugli endpoint VPC per le azioni di Lake Formation

Il seguente esempio di politica degli endpoint VPC per Lake Formation consente la vendita di credenziali utilizzando le autorizzazioni di Lake Formation. Puoi utilizzare questa policy per eseguire query utilizzando le autorizzazioni di Lake Formation da un cluster Amazon Redshift o da un cluster situato in Amazon EMR una sottorete privata.

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

> **Note**
>
> Se non alleghi una policy quando crei un endpoint, viene allegata una policy predefinita che consente l'accesso completo al servizio.

Per ulteriori informazioni, consulta questi argomenti nella documentazione di Amazon VPC:

- Che cos'è Amazon VPC?
- Crea un endpoint di interfaccia
- Usa le policy degli endpoint VPC

Tutorial

I seguenti tutorial sono organizzati in tre percorsi e forniscono step-by-step istruzioni su come creare un data lake, importare dati, condividere e proteggere i data lake utilizzando: AWS Lake Formation

1. Crea un data lake e inserisci dati: impara a creare un data lake e usa i blueprint per spostare, archiviare, catalogare, pulire e organizzare i dati. Imparerai anche a configurare tabelle gestite. Una tabella governata è un nuovo tipo di tabella Amazon S3 che supporta transazioni atomiche, consistenti, isolate e durevoli (ACID).

   Prima di iniziare, assicurati di aver completato i passaggi indicati. Guida introduttiva a Lake Formation

   - Creazione di un data lake da un' AWS CloudTrail origine

     Crea e carica il tuo primo data lake utilizzando i tuoi CloudTrail log come fonte di dati.

   - Creazione di un data lake da una sorgente JDBC in Lake Formation

     Crea un data lake utilizzando uno dei tuoi data store accessibili con JDBC, ad esempio un database relazionale, come origine dati.

2. Protezione dei data lake: impara a utilizzare controlli di accesso basati su tag e a livello di riga per proteggere e gestire in modo efficace l'accesso ai tuoi data lake.

   - Impostazione delle autorizzazioni per i formati di archiviazione a tabella aperta in Lake Formation

     Questo tutorial dimostra come impostare le autorizzazioni per i formati di tabelle transazionali open source (tabelle Apache Iceberg, Apache Hudi e Linux Foundation Delta Lake) in Lake Formation.

   - Gestione di un data lake utilizzando il controllo degli accessi basato su tag Lake Formation

     Impara a gestire l'accesso ai dati all'interno di un data lake utilizzando il controllo degli accessi basato su tag in Lake Formation.

   - Protezione dei data lake con il controllo degli accessi a livello di riga

     Scopri come configurare le autorizzazioni a livello di riga che ti consentono di limitare l'accesso a righe specifiche in base alle politiche di conformità e governance dei dati in Lake Formation.

3. Condivisione dei dati: impara a condividere in modo sicuro i tuoi dati Account AWS utilizzando il controllo degli accessi basato su tag (TBAC) e gestisci le autorizzazioni granulari sui set di dati condivisi tra. Account AWS
   - Condivisione di un data lake utilizzando il controllo degli accessi basato su tag Lake Formation e risorse denominate

     In questo tutorial, imparerai come condividere in modo sicuro i tuoi dati Account AWS utilizzando Lake Formation.
   - Condivisione di un data lake utilizzando il controllo granulare degli accessi di un data lake

     In questo tutorial, imparerai come condividere in modo rapido e semplice set di dati utilizzando Lake Formation quando gestisci più set Account AWS conAWS Organizations.

Argomenti

- Creazione di un data lake da un' AWS CloudTrail origine
- Creazione di un data lake da una sorgente JDBC in Lake Formation
- Impostazione delle autorizzazioni per i formati di archiviazione a tabella aperta in Lake Formation
- Gestione di un data lake utilizzando il controllo degli accessi basato su tag Lake Formation
- Protezione dei data lake con il controllo degli accessi a livello di riga
- Condivisione di un data lake utilizzando il controllo degli accessi basato su tag Lake Formation e risorse denominate
- Condivisione di un data lake utilizzando il controllo granulare degli accessi di un data lake

Creazione di un data lake da un' AWS CloudTrail origine

Questo tutorial ti guida attraverso le azioni da intraprendere sulla console Lake Formation per creare e caricare il tuo primo data lake da una AWS CloudTrail fonte.

Passaggi di alto livello per la creazione di un data lake

1. Registra un percorso Amazon Simple Storage Service (Amazon S3) come data lake.
2. Concedi a Lake Formation le autorizzazioni per scrivere nel Data Catalog e nelle posizioni Amazon S3 nel data lake.
3. Crea un database per organizzare le tabelle di metadati nel Data Catalog.

4. Usa un blueprint per creare un flusso di lavoro. Esegui il flusso di lavoro per importare dati da una fonte di dati.
5. Configura le tue autorizzazioni Lake Formation per consentire ad altri di gestire i dati nel Data Catalog e nel data lake.
6. Configura Amazon Athena per interrogare i dati che hai importato nel tuo data lake Amazon S3.
7. Per alcuni tipi di data store, configura Amazon Redshift Spectrum per interrogare i dati che hai importato nel tuo data lake Amazon S3.

Argomenti

Destinatari principali

La tabella seguente elenca i ruoli utilizzati in questo tutorial per creare un data lake.

Destinatari principali

Ruolo	Descrizione
Amministratore IAM	Ha la politica AWS gestita:`AdministratorAccess` . Può creare ruoli IAM e bucket Amazon S3.
Amministratore del data lake	Utente che può accedere al catalogo dati, creare database e concedere autorizzazioni Lake Formation ad altri utenti. Dispone di meno autorizzazioni IAM rispetto all'amministratore IAM, ma sufficienti per amministrare il data lake.
Analista dei dati	Utente che può eseguire query sul data lake. Dispone solo delle autorizzazioni sufficienti per eseguire le query.
Ruolo del workflow	Ruolo con le politiche IAM richieste per eseguire un flusso di lavoro. Per ulteriori informazioni, consulta (Facoltativo) Crea un ruolo IAM per i flussi di lavoro.

Prerequisiti

Prima di iniziare:

- Assicurati di aver completato le attività inConfigurazione di AWS Lake Formation.
- Conosci la posizione dei tuoi CloudTrail registri.
- Athena richiede all'analista di dati di creare un bucket Amazon S3 per archiviare i risultati delle query prima di utilizzare Athena.

Si presume la familiarità con AWS Identity and Access Management (IAM). Per informazioni su IAM, consulta la IAM User Guide.

Fase 1: Creare un utente analista di dati

Questo utente dispone del set minimo di autorizzazioni per interrogare il data lake.

1. Aprire la console IAM all'indirizzo https://console.aws.amazon.com/iam. Accedi come utente amministratore che hai creato Creazione di un utente amministratore o come utente con la policy `AdministratorAccess` AWS gestita.
2. Crea un utente denominato `datalake_user` con le seguenti impostazioni:
   - Abilita AWS Management Console l'accesso.
   - Imposta una password e non richiedi la reimpostazione della password.
   - Allega la politica `AmazonAthenaFullAccess` AWS gestita.
   - Allega la seguente politica in linea. Assegnare un nome alla policy `DatalakeUserBasic`.

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

Passaggio 2: Aggiungere le autorizzazioni per leggere i AWS CloudTrail registri al ruolo del flusso di lavoro

1. Allega la seguente politica in linea al ruolo. `LakeFormationWorkflowRole` La policy concede il permesso di leggere i tuoi AWS CloudTrail log. Assegnare un nome alla policy `DatalakeGetCloudTrail`.

   Per creare il ruolo `LakeFormationWorkflowRole`, consulta (Facoltativo) Crea un ruolo IAM per i flussi di lavoro.

   > **Important**
   >
   > Sostituisci <your-s3-cloudtrail-bucket>con la posizione dei tuoi CloudTrail dati in Amazon S3.

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

2. Verifica che al ruolo siano associate tre policy.

Fase 3: creare un bucket Amazon S3 per il data lake

Crea il bucket Amazon S3 che sarà la posizione principale del tuo data lake.

1. Apri la console Amazon S3 all'indirizzo https://console.aws.amazon.com/s3/ e accedi come utente amministratore con cui hai creato. Creazione di un utente amministratore
2. Scegli Crea bucket e segui la procedura guidata per creare un bucket denominato*<yourName>*-`datalake-cloudtrail`, dove sono <yourName>il tuo nome e cognome. Ad esempio: `jdoe-datalake-cloudtrail`.

Per istruzioni dettagliate sulla creazione di un bucket Amazon S3, consulta Creazione di un bucket.

Fase 4: Registrare un percorso Amazon S3

Registra un percorso Amazon S3 come posizione principale del tuo data lake.

1. Aprire la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/. Accedi come amministratore del data lake.
2. Nel riquadro di navigazione, in Registra e inserisci, scegli Data lake locations.
3. Scegli Registra posizione e poi Sfoglia.
4. Seleziona il *<yourName>*`-datalake-cloudtrail` bucket che hai creato in precedenza, accetta il ruolo IAM predefinito`AWSServiceRoleForLakeFormationDataAccess`, quindi scegli Registra posizione.

   Per ulteriori informazioni sulla registrazione delle sedi, consulta. Aggiungere una posizione Amazon S3 al tuo data lake

Passaggio 5: concedere le autorizzazioni per la localizzazione dei dati

I responsabili devono disporre delle autorizzazioni di localizzazione dei dati su una posizione di data lake per creare tabelle o database di Data Catalog che puntano a tale posizione. È necessario concedere le autorizzazioni per la localizzazione dei dati al ruolo IAM per i flussi di lavoro in modo che il flusso di lavoro possa scrivere nella destinazione di inserimento dei dati.

1. Nel riquadro di navigazione, in Autorizzazioni, scegli Posizioni dei dati.
2. Scegli Concedi e, nella finestra di dialogo Concedi autorizzazioni, effettua le seguenti selezioni:
   a. Per utenti e ruoli IAM, scegli. `LakeFormationWorkflowRole`
   b. Per le posizioni di archiviazione, scegli il tuo *<yourName>*`-datalake-cloudtrail` bucket.
3. Scegli Concessione.

Per ulteriori informazioni sulle autorizzazioni per la localizzazione dei dati, consulta. Underlying data access control

Fase 6: Creare un database nel Data Catalog

Le tabelle di metadati nel Lake Formation Data Catalog sono archiviate all'interno di un database.

1. Nel riquadro di navigazione, in Catalogo dati, scegli Database.
2. Scegli Crea database e in Dettagli del database, inserisci il nome`lakeformation_cloudtrail`.
3. Lascia vuoti gli altri campi e scegli Crea database.

Passaggio 7: concedere le autorizzazioni per i dati

È necessario concedere le autorizzazioni per creare tabelle di metadati nel Catalogo dati. Poiché il flusso di lavoro verrà eseguito con il ruolo`LakeFormationWorkflowRole`, è necessario concedere queste autorizzazioni al ruolo.

1. Nella console Lake Formation, nel riquadro di navigazione, in Catalogo dati, scegli Databases.
2. Scegli il `lakeformation_cloudtrail` database, quindi, dall'elenco a discesa Azioni, scegli Concedi sotto la voce Autorizzazioni.
3. Nella finestra di dialogo Concedi le autorizzazioni per i dati, effettua le seguenti selezioni:

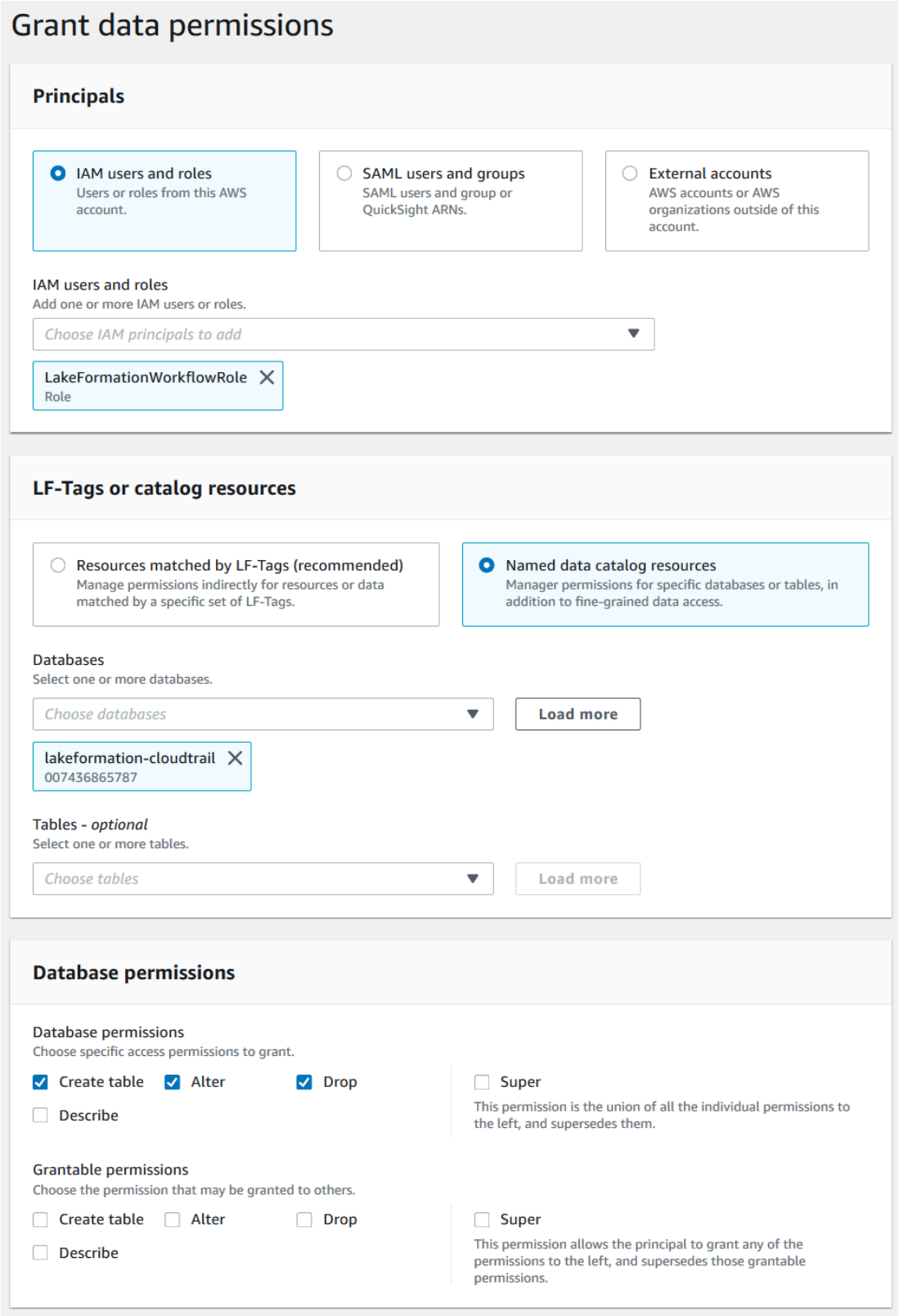
   a. In Principali, per utenti e ruoli IAM, scegli. `LakeFormationWorkflowRole`

   b. In LF-Tags o Catalog resources, scegli Named data catalog resources.

   c. Per i database, dovresti vedere che il `lakeformation_cloudtrail` database è già stato aggiunto.

   d. In Autorizzazioni del database, seleziona Crea tabella, Alter e Drop e deseleziona Super se è selezionato.

   La finestra di dialogo Concedi le autorizzazioni per i dati ora dovrebbe avere l'aspetto di questa schermata.



4. Scegli Concessione.

Per ulteriori informazioni sulla concessione delle autorizzazioni di Lake Formation, vedere. Gestione delle autorizzazioni di Lake Formation

Fase 8: Utilizzare un blueprint per creare un flusso di lavoro

Per leggere CloudTrail i log, comprenderne la struttura, creare le tabelle appropriate nel Data Catalog, dobbiamo impostare un flusso di lavoro composto da AWS Glue crawler, job, trigger e workflow. I progetti di Lake Formation semplificano questo processo.

Il flusso di lavoro genera i job, i crawler e i trigger che rilevano e inseriscono i dati nel tuo data lake. Crei un flusso di lavoro basato su uno dei blueprint predefiniti di Lake Formation.

1. Nella console di Lake Formation, nel pannello di navigazione, scegli Blueprints, quindi scegli Usa blueprint.
2. Nella pagina Usa un blueprint, in Tipo di blueprint, scegli. AWS CloudTrail
3. In Importa fonte, scegli una CloudTrail fonte e una data di inizio.
4. In Import target, specifica questi parametri:

Database di destinazione	`lakeformation_cloudtrail`
Posizione di archiviazione di destinazione	`s3://<yourName> -datalake-cloudtrail`
Formato dei dati	Parquet

5. Per la frequenza di importazione, scegli Esegui su richiesta.
6. In Opzioni di importazione, specificate questi parametri:

Nome del flusso di lavoro	`lakeformationcloudtrailtest`
Ruolo IAM	`LakeFormationWorkflowRole`
Prefisso della tabella	`cloudtrailtest`

> **Note**
>
> Deve essere in minuscolo.

7. Scegli Crea e attendi che la console segnali che il flusso di lavoro è stato creato correttamente.

> **Tip**
>
> Hai ricevuto il seguente messaggio di errore?
> `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...`
> In tal caso, verifica di aver sostituito <account-id>nella politica in linea per l'utente amministratore del data lake un numero di AWS account valido.

Passaggio 9: Esegui il flusso di lavoro

Poiché hai specificato che il flusso di lavoro è run-on-demand, devi avviarlo manualmente.

- Nella pagina Blueprint, seleziona il flusso di lavoro `lakeformationcloudtrailtest` e nel menu Azioni scegli Avvia.

  Durante l'esecuzione del flusso di lavoro, è possibile visualizzarne l'avanzamento nella colonna Stato dell'ultima esecuzione. Scegli il pulsante di aggiornamento di tanto in tanto.

  Lo stato va da RUNNING, a Discovering, a Importing, a COMPLETED.

  Al termine del flusso di lavoro:

  - Il Data Catalog avrà nuove tabelle di metadati.
  - CloudTrail I log verranno inseriti nel data lake.

  Se il flusso di lavoro fallisce, procedi come segue:

  a. Seleziona il flusso di lavoro e nel menu Azioni scegli Visualizza grafico.

     Il flusso di lavoro viene aperto nella AWS Glue console.

  b. Verifica che il flusso di lavoro sia selezionato e scegli la scheda History (Cronologia).

  c. In Cronologia, seleziona l'esecuzione più recente e scegli Visualizza i dettagli dell'esecuzione.

d. Seleziona un processo o un crawler non riuscito nel grafico dinamico (di runtime) ed esamina il messaggio di errore. I nodi con errori sono rossi o gialli.

Fase 10: concedere SELECT sui tavoli

È necessario concedere l'`SELECT`autorizzazione per le nuove tabelle del Catalogo dati in modo che l'analista dei dati possa interrogare i dati a cui fanno riferimento le tabelle.

> **Note**
>
> Un flusso di lavoro concede automaticamente l'`SELECT`autorizzazione per le tabelle che crea all'utente che lo ha eseguito. Poiché l'amministratore del data lake ha eseguito questo flusso di lavoro, è necessario concederlo `SELECT` all'analista dei dati.

1. Nella console Lake Formation, nel riquadro di navigazione, in Catalogo dati, scegli Databases.
2. Scegli il `lakeformation_cloudtrail` database, quindi, dall'elenco a discesa Azioni, scegli Concedi sotto la voce Autorizzazioni.
3. Nella finestra di dialogo Concedi le autorizzazioni per i dati, effettua le seguenti selezioni:
   a. In Principali, per utenti e ruoli IAM, scegli. `datalake_user`
   b. In LF-Tags o Catalog resources, scegli Named data catalog resources.
   c. Per Databases, il `lakeformation_cloudtrail` database dovrebbe essere già selezionato.
   d. Per Tabelle, scegliete`cloudtrail`il `test-cloudtrail`.
   e. In Autorizzazioni per tabelle e colonne, scegli Seleziona.
4. Scegli Concessione.

Il passaggio successivo viene eseguito come analista dei dati.

Passaggio 11: interrogare il data lake utilizzando Amazon Athena

Usa la Amazon Athena console per interrogare i CloudTrail dati nel tuo data lake.

1. Apri la console Athena all'indirizzo https://console.aws.amazon.com/athena/ e accedi come analista di dati, utente. `datalake_user`

2. Se necessario, scegli Inizia per passare all'editor di query Athena.
3. Per Origine dati scegliere AwsDataCatalog.
4. Per Database, scegliere `lakeformation_cloudtrail`.

   L'elenco delle tabelle viene compilato.
5. Nel menu a discesa (3 punti disposti orizzontalmente) accanto alla tabella**`cloudtrailtest-cloudtrail`**, scegliete Anteprima tabella, quindi scegliete Esegui.

   La query viene eseguita e visualizza 10 righe di dati.

   Se non hai mai usato Athena prima, devi prima configurare una posizione Amazon S3 nella console Athena per archiviare i risultati delle query. `datalake_user`È necessario disporre delle autorizzazioni necessarie per accedere al bucket Amazon S3 scelto.

> **Note**
>
> Ora che hai completato il tutorial, concedi le autorizzazioni per i dati e le autorizzazioni per la localizzazione dei dati ai responsabili della tua organizzazione.

Creazione di un data lake da una sorgente JDBC in Lake Formation

Questo tutorial illustra i passaggi da eseguire sulla AWS Lake Formation console per creare e caricare il primo data lake da una sorgente JDBC utilizzando Lake Formation.

Argomenti

- Fase 8: Utilizzare un blueprint per creare un flusso di lavoro
- Passaggio 9: Esegui il flusso di lavoro
- Fase 10: concedere SELECT sui tavoli
- Passaggio 11: interrogare il data lake utilizzando Amazon Athena
- Passaggio 12: interroga i dati nel data lake utilizzando Amazon Redshift Spectrum
- Fase 13: concedere o revocare le autorizzazioni di Lake Formation utilizzando Amazon Redshift Spectrum

Destinatari principali

La tabella seguente elenca i ruoli utilizzati in questo tutorial AWS Lake FormationJDBC.

Ruolo	Descrizione
Amministratore IAM	Un utente che può creare utenti e ruoli AWS Identity and Access Management (IAM) e bucket Amazon Simple Storage Service (Amazon S3). Dispone della politica `AdministratorAccess` AWS gestita.
Amministratore del data lake	Un utente che può accedere al Data Catalog, creare database e concedere autorizzazioni Lake Formation ad altri utenti. Dispone di meno autorizzazioni IAM rispetto all'amministratore IAM, ma sufficienti per amministrare il data lake.
Analista dei dati	Un utente che può eseguire query sul data lake. Dispone solo delle autorizzazioni sufficienti per eseguire le query.
Ruolo del workflow	Un ruolo con le politiche IAM richieste per eseguire un flusso di lavoro.

Per informazioni sui prerequisiti per il completamento del tutorial, consultaPrerequisiti del tutorial JDBC.

Prerequisiti del tutorial JDBC

Prima di iniziare il tutorial su AWS Lake Formation JDBC, assicuratevi di aver fatto quanto segue:

- Completare le operazioni descritte in Guida introduttiva a Lake Formation.
- Scegli un archivio dati accessibile con JDBC che desideri utilizzare per il tutorial.
- Raccogli le informazioni necessarie per creare una AWS Glue connessione di tipo JDBC. Questo oggetto Data Catalog include l'URL del data store, le credenziali di accesso e, se il data store è stato creato in un Amazon Virtual Private Cloud (Amazon VPC), informazioni di configurazione aggiuntive specifiche per VPC. Per ulteriori informazioni, consulta Definizione delle connessioni nel catalogo AWS Glue dati nella Guida per gli sviluppatori. AWS Glue

Il tutorial presuppone che tu abbia familiarità con AWS Identity and Access Management (IAM). Per informazioni su IAM, consulta la IAM User Guide.

Per iniziare, procedi athe section called “Fase 1: Creare un utente analista di dati”.

Fase 1: Creare un utente analista di dati

In questo passaggio, crei un utente AWS Identity and Access Management (IAM) che funga da analista di dati per il tuo data lake in. AWS Lake Formation

Questo utente dispone del set minimo di autorizzazioni per interrogare il data lake.

1. Aprire la console IAM all'indirizzo https://console.aws.amazon.com/iam. Accedi come utente amministratore che hai creato Creazione di un utente amministratore o come utente con la policy `AdministratorAccess` AWS gestita.
2. Crea un utente denominato `datalake_user` con le seguenti impostazioni:
   - Abilita AWS Management Console l'accesso.
   - Imposta una password e non richiedi la reimpostazione della password.
   - Allega la politica `AmazonAthenaFullAccess` AWS gestita.
   - Allega la seguente politica in linea. Assegnare un nome alla policy `DatalakeUserBasic`.

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

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

Fase 2: Creare una connessione in AWS Glue

> **Note**
>
> Salta questo passaggio se disponi già di una AWS Glue connessione alla tua fonte di dati JDBC.

AWS Lake Formationaccede alle sorgenti dati JDBC tramite una connessione. AWS Glue Una connessione è un oggetto Data Catalog che contiene tutte le informazioni necessarie per connettersi all'origine dati. È possibile creare una connessione utilizzando la AWS Glue console.

Per creare una connessione

1. Apri AWS Glue la console all'indirizzo https://console.aws.amazon.com/glue/ e accedi come utente amministratore con cui hai creato il fileCreazione di un utente amministratore.
2. Nel riquadro di navigazione, in Data catalog (Catalogo dati), seleziona Connections (Connessioni).

3. Nella pagina Connectors (Connettori), seleziona Create custom connector (Crea connettore personalizzato).
4. Nella pagina delle proprietà del connettore, immettete **`datalake-tutorial`** come nome della connessione e scegliete JDBC come tipo di connessione. Quindi scegli Next (Successivo).
5. Continuate con la procedura guidata di connessione e salvate la connessione.

   Per informazioni sulla creazione di una connessione, consulta le proprietà della connessione AWS Glue JDBC nella Guida per gli AWS Glue sviluppatori.

Fase 3: creare un bucket Amazon S3 per il data lake

In questo passaggio, crei il bucket Amazon Simple Storage Service (Amazon S3) che deve essere la posizione principale del tuo data lake.

1. Apri la console Amazon S3 all'indirizzo https://console.aws.amazon.com/s3/ e accedi come utente amministratore con cui hai creato. Creazione di un utente amministratore
2. Scegli Crea bucket e segui la procedura guidata per creare un bucket denominato*<yourName>*-`datalake-tutorial`, dove sono <yourName>il tuo nome e cognome. Ad esempio: `jdoe-datalake-tutorial`.

   Per istruzioni dettagliate sulla creazione di un bucket Amazon S3, vedi Come si crea un bucket S3? nella Guida per l'utente di Amazon Simple Storage Service.

Fase 4: Registrare un percorso Amazon S3

In questa fase, registri un percorso Amazon Simple Storage Service (Amazon S3) come posizione principale del tuo data lake.

1. Aprire la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/. Accedi come amministratore del data lake.
2. Nel riquadro di navigazione, in Registra e inserisci, scegli Data lake locations.
3. Scegli Registra posizione, quindi scegli Sfoglia.
4. Seleziona il *<yourName>*`-datalake-tutorial` bucket che hai creato in precedenza, accetta il ruolo IAM predefinito`AWSServiceRoleForLakeFormationDataAccess`, quindi scegli Registra posizione.

Per ulteriori informazioni sulla registrazione delle sedi, consulta. Aggiungere una posizione Amazon S3 al tuo data lake

Passaggio 5: concedere le autorizzazioni per la localizzazione dei dati

I responsabili devono disporre delle autorizzazioni di localizzazione dei dati su una posizione di data lake per creare tabelle o database di Data Catalog che puntano a quella posizione. È necessario concedere le autorizzazioni per la localizzazione dei dati al ruolo IAM per i flussi di lavoro in modo che il flusso di lavoro possa scrivere nella destinazione di inserimento dei dati.

1. Sulla console Lake Formation, nel riquadro di navigazione, in Autorizzazioni, scegli Posizioni dati.
2. Scegliete Concedi e, nella finestra di dialogo Concedi autorizzazioni, effettuate le seguenti operazioni:
   a. Per utenti e ruoli IAM, scegli`LakeFormationWorkflowRole`.
   b. Per le posizioni di archiviazione, scegli il tuo `<yourName>-datalake-tutorial` bucket.
3. Scegli Concessione.

Per ulteriori informazioni sulle autorizzazioni per la localizzazione dei dati, consulta. Underlying data access control

Fase 6: Creare un database nel Data Catalog

Le tabelle di metadati nel Lake Formation Data Catalog sono archiviate all'interno di un database.

1. Nella console Lake Formation, nel riquadro di navigazione, in Catalogo dati, scegli Databases.
2. Scegli Crea database e in Dettagli del database, inserisci il nome`lakeformation_tutorial`.
3. Lascia vuoti gli altri campi e scegli Crea database.

Passaggio 7: concedere le autorizzazioni per i dati

È necessario concedere le autorizzazioni per creare tabelle di metadati nel Data Catalog. Poiché il flusso di lavoro viene eseguito con il ruolo`LakeFormationWorkflowRole`, è necessario concedere queste autorizzazioni al ruolo.

1. Nella console Lake Formation, nel riquadro di navigazione, in Autorizzazioni, scegli Autorizzazioni Data lake.
2. Scegli Concedi e nella finestra di dialogo Concedi le autorizzazioni per i dati, procedi come segue:
   a. In Principal, per utenti e ruoli IAM, scegli. `LakeFormationWorkflowRole`
   b. In LF-Tags o Catalog resources, scegli Named data catalog resources.
   c. Per Database, scegli il database che hai creato in precedenza,. `lakeformation_tutorial`
   d. In Autorizzazioni del database, seleziona Crea tabella, Alter e Drop e deseleziona Super se è selezionato.
3. Scegli Concessione.

Per ulteriori informazioni sulla concessione delle autorizzazioni di Lake Formation, vedere. Panoramica delle autorizzazioni di Lake Formation

Fase 8: Utilizzare un blueprint per creare un flusso di lavoro

Il AWS Lake Formation flusso di lavoro genera i AWS Glue job, i crawler e i trigger che rilevano e inseriscono dati nel tuo data lake. Crei un flusso di lavoro basato su uno dei blueprint predefiniti di Lake Formation.

1. Nella console Lake Formation, nel pannello di navigazione, scegli Blueprints, quindi scegli Usa blueprint.
2. Nella pagina Usa un blueprint, in Tipo di blueprint, scegli Database snapshot.
3. In Origine di importazione, per Connessione al database, scegli la connessione che hai appena creato o scegli una connessione esistente per la tua origine dati. `datalake-tutorial`
4. Per Percorso dei dati di origine, inserisci nel modulo *`<database>`*/*`<schema>`*/*`<table>`* il percorso da cui importare i dati.

   È possibile sostituire lo schema o la tabella con il carattere jolly percentuale (%). <schema><database>Per i database che supportano gli schemi, inserisci<database>/<schema>/% per abbinare tutte le tabelle contenute all'interno. <database>Oracle Database e MySQL non supportano lo schema nel percorso; inserisci invece /%. Per Oracle Database, <database>è l'identificatore di sistema (SID).

Ad esempio, se un database Oracle ha `orcl` come SID, immettilo in modo che `orcl/%` corrisponda a tutte le tabelle a cui ha accesso l'utente specificato nella connessione JDCB.

> **Important**
>
> Questo campo fa distinzione tra minuscole e maiuscole.

5. In Import target, specifica questi parametri:

Database di destinazione	`lakeformation_tutorial`
Posizione di archiviazione di destinazione	`s3://<yourName> -datalake-tutorial`
Formato dei dati	(Scegli Parquet o CSV)

6. Per la frequenza di importazione, scegli Esegui su richiesta.
7. In Opzioni di importazione, specificate questi parametri:

Nome del flusso di lavoro	`lakeformationjdbctest`
Ruolo IAM	`LakeFormationWorkflowRole`
Prefisso della tabella	`jdbctest` **Note** Deve essere in minuscolo.

8. Scegli Crea e attendi che la console segnali che il flusso di lavoro è stato creato correttamente.

> **Tip**
>
> Hai ricevuto il seguente messaggio di errore?
> `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...
In tal caso, verifica di aver sostituito <account-id>nella politica in linea per l'utente amministratore del data lake un numero di AWS account valido.

Passaggio 9: Esegui il flusso di lavoro

Poiché hai specificato che il flusso di lavoro è run-on-demand, devi avviarlo manualmente inAWS Lake Formation.

1. Nella console Lake Formation, nella pagina Blueprints, seleziona il flusso di lavoro`lakeformationjdbctest`.
2. Scegli Azioni, quindi scegli Avvia.
3. Durante l'esecuzione del flusso di lavoro, visualizzane l'avanzamento nella colonna Stato dell'ultima esecuzione. Scegli il pulsante di aggiornamento di tanto in tanto.

   Lo stato va da RUNNING, a Discovering, a Importing, a COMPLETED.

   Quando il flusso di lavoro è completo:

   - Il Data Catalog ha nuove tabelle di metadati.
   - I tuoi dati vengono inseriti nel data lake.

   Se il flusso di lavoro fallisce, procedi come segue:

   a. Seleziona un flusso di lavoro. Scegliete Azioni, quindi scegliete Visualizza grafico.

      Il flusso di lavoro si apre nella AWS Glue console.

   b. Seleziona il flusso di lavoro e scegli la scheda Cronologia.

   c. Seleziona l'esecuzione più recente e scegli Visualizza i dettagli della corsa.

   d. Seleziona un processo o un crawler non riuscito nel grafico dinamico (runtime) ed esamina il messaggio di errore. I nodi con errori sono rossi o gialli.

Fase 10: concedere SELECT sui tavoli

È necessario concedere l'`SELECT`autorizzazione per le nuove tabelle del Data Catalog in AWS Lake Formation modo che l'analista dei dati possa interrogare i dati a cui puntano le tabelle.

> **Note**
>
> Un flusso di lavoro concede automaticamente l'`SELECT`autorizzazione per le tabelle che crea all'utente che lo ha eseguito. Poiché l'amministratore del data lake ha eseguito questo flusso di lavoro, è necessario concederlo `SELECT` all'analista dei dati.

1. Nella console Lake Formation, nel riquadro di navigazione, in Autorizzazioni, scegli Autorizzazioni Data lake.
2. Scegli Concedi e nella finestra di dialogo Concedi le autorizzazioni per i dati, procedi come segue:
   a. In Principal, per utenti e ruoli IAM, scegli. `datalake_user`
   b. In LF-Tags o Catalog resources, scegli Named data catalog resources.
   c. Per Database, scegliete. `lakeformation_tutorial`

      Viene compilato l'elenco Tabelle.
   d. Per Tabelle, scegli una o più tabelle dalla tua fonte di dati.
   e. In Autorizzazioni per tabelle e colonne, scegli Seleziona.
3. Scegli Concessione.

Il passaggio successivo viene eseguito come analista dei dati.

Passaggio 11: interrogare il data lake utilizzando Amazon Athena

Usa la Amazon Athena console per interrogare i dati nel tuo data lake.

1. Apri la console Athena all'indirizzo https://console.aws.amazon.com/athena/ e accedi come analista di dati, utente. `datalake_user`
2. Se necessario, scegli Inizia per passare all'editor di query Athena.
3. Per Origine dati scegliere AwsDataCatalog.
4. Per Database, scegliere `lakeformation_tutorial`.

L'elenco delle tabelle viene compilato.

5. Nel menu pop-up accanto a una delle tabelle, scegli Anteprima tabella.

   La query viene eseguita e visualizza 10 righe di dati.

Passaggio 12: interroga i dati nel data lake utilizzando Amazon Redshift Spectrum

Puoi configurare Amazon Redshift Spectrum per interrogare i dati che hai importato nel tuo data lake Amazon Simple Storage Service (Amazon S3). Innanzitutto, crea un ruolo AWS Identity and Access Management (IAM) da utilizzare per avviare il cluster Amazon Redshift e interrogare i dati di Amazon S3. Quindi, concedi a questo ruolo le `Select` autorizzazioni sulle tabelle che desideri interrogare. Quindi, concedi all'utente le autorizzazioni per utilizzare l'editor di query di Amazon Redshift. Infine, crea un cluster Amazon Redshift ed esegui le query.

Crei il cluster come amministratore e interroghi il cluster come analista di dati.

Per ulteriori informazioni su Amazon Redshift Spectrum, consulta Using Amazon Redshift Spectrum to External Data nella Amazon Redshift Database Developer Guide.

Per configurare le autorizzazioni per eseguire query Amazon Redshift

1. Aprire la console IAM all'indirizzo https://console.aws.amazon.com/iam/. Accedi come utente amministratore in cui hai creato Creazione di un utente amministratore (nome utente`Administrator`) o come utente con la `AdministratorAccess` AWS policy gestita.
2. Nel pannello di navigazione, selezionare Policy.

   Se è la prima volta che si seleziona Policies (Policy), verrà visualizzata la pagina Welcome to Managed Policies (Benvenuto nelle policy gestite). Seleziona Get Started (Inizia).
3. Scegli Create Policy (Crea policy).
4. Scegliere la scheda JSON.
5. Incolla il seguente documento di policy 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. Una volta terminato, selezionare Review (Rivedi) per rivedere la policy. In Validatore di policy vengono segnalati eventuali errori di sintassi.
7. Nella pagina Rivedi la politica, inserisci **`RedshiftLakeFormationPolicy`** il nome della politica che stai creando. (Opzionale) Immettere una Description (descrizione). Consulta il Summary (Riepilogo) della policy per visualizzare le autorizzazioni concesse dalla policy. Seleziona Create policy (Crea policy) per salvare il proprio lavoro.
8. Nel pannello di navigazione della console IAM, scegliere Ruoli e quindi Crea ruolo.
9. In Seleziona tipo di entità attendibile, scegli Servizio AWS.
10. Scegliere il servizio Amazon Redshift per assumere questo ruolo.
11. Selezionare il caso d'uso Redshift Customizable (Redshift personalizzabile) per il servizio. Quindi scegliere Next: Permissions (Successivo: Autorizzazioni).
12. Cerca la politica di autorizzazione che hai creato e seleziona la casella di controllo accanto al nome della politica nell'elenco. `RedshiftLakeFormationPolicy`
13. Scegliere Next: Tags (Successivo: Tag).
14. Scegliere Next:Review (Successivo: Rivedi).
15. In Role name (Nome ruolo), inserire il nome **`RedshiftLakeFormationRole`**.
16. (Facoltativo) In Role description (Descrizione ruolo), immettere una descrizione per il nuovo ruolo.

17. Verificare il ruolo e quindi scegliere Create role (Crea ruolo).

Per concedere **`Select`** le autorizzazioni sulla tabella da interrogare nel database Lake Formation

1. Aprire la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/. Accedi come amministratore del data lake.
2. Nel riquadro di navigazione, in Autorizzazioni, scegli Autorizzazioni Data lake, quindi scegli Concedi.
3. Inserisci le informazioni che seguono:
   - Per gli utenti e i ruoli IAM, scegli il ruolo IAM che hai creato,. `RedshiftLakeFormationRole` Quando si esegue l'editori di query di Amazon Redshift, utilizzare il ruolo IAM per le autorizzazioni ai dati.
   - Per Database, scegliere `lakeformation_tutorial`.

     L'elenco delle tabelle viene compilato.
   - Per Tabella, scegli una tabella all'interno della fonte di dati da interrogare.
   - Scegli l'autorizzazione Seleziona tabella.
4. Scegli Concessione.

Per configurare Amazon Redshift Spectrum ed eseguire query

1. Apri la console Amazon Redshift all'indirizzo. https://console.aws.amazon.com/redshift Accedi come utente`Administrator`.
2. Scegli Crea cluster.
3. Nella pagina Crea cluster, inserisci `redshift-lakeformation-demo` l'identificatore del cluster.
4. Per il tipo di nodo, seleziona dc2.large.
5. Scorri verso il basso e, in Configurazioni del database, inserisci o accetta questi parametri:
   - Nome utente amministratore: `awsuser`
   - Password utente amministratore: (*`Choose a password`*)
6. Espandi le autorizzazioni del cluster e, per i ruoli IAM disponibili, scegli RedshiftLakeFormationRole. Scegli Add IAM role (Aggiungi ruolo IAM).

7. Se devi utilizzare una porta diversa dal valore predefinito 5439, accanto a Configurazioni aggiuntive, disattiva l'opzione Usa valori predefiniti. Espandi la sezione relativa alle configurazioni del database e inserisci un nuovo numero di porta del database.
8. Scegli Crea cluster.

   Viene caricata la pagina Cluster.
9. Attendi che lo stato del cluster diventi Disponibile. Scegli periodicamente l'icona di aggiornamento.
10. Concedi all'analista dei dati l'autorizzazione a eseguire query sul cluster. Per fare ciò, completare questa procedura:

    a. Apri la console IAM all'indirizzo https://console.aws.amazon.com/iam/ e accedi come utente. `Administrator`

    b. Nel riquadro di navigazione, scegli Utenti e allega all'utente le seguenti politiche gestite`datalake_user`.

    - `AmazonRedshiftQueryEditor`
    - `AmazonRedshiftReadOnlyAccess`
11. Esci dalla console Amazon Redshift e accedi nuovamente come utente. `datalake_user`
12. Nella barra degli strumenti verticale a sinistra, scegli l'icona EDITOR per aprire l'editor di query e connetterti al cluster. Se viene visualizzata la finestra di dialogo Connetti al database, scegliete il nome `redshift-lakeformation-demo` del cluster e immettete il nome **`dev`** del database**`awsuser`**, il nome utente e la password che avete creato. Quindi scegli Connect to database (Collegati al database).

    > **Note**
    >
    > Se non ti vengono richiesti i parametri di connessione e nell'editor di query è già selezionato un altro cluster, scegli Cambia connessione per aprire la finestra di dialogo Connetti al database.
13. Nella casella di testo New Query 1, inserisci ed esegui la seguente istruzione per mappare il database `lakeformation_tutorial` in Lake Formation al nome dello schema Amazon Redshift: `redshift_jdbc`

> **Important**
>
> <account-id>Sostituiscilo con un numero di AWS account valido e <region>con un nome di AWS regione valido (ad esempio,`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. Nell'elenco degli schemi in Seleziona schema, scegli redshift_jdbc.

    L'elenco delle tabelle viene compilato. L'editor di query mostra solo le tabelle per le quali ti sono state concesse le autorizzazioni per il data lake di Lake Formation.

15. Nel menu a comparsa accanto al nome di una tabella, scegli Anteprima dati.

    Amazon Redshift restituisce le prime 10 righe.

    Ora puoi eseguire query sulle tabelle e sulle colonne per le quali disponi delle autorizzazioni.

Fase 13: concedere o revocare le autorizzazioni di Lake Formation utilizzando Amazon Redshift Spectrum

Amazon Redshift supporta la possibilità di concedere e revocare le autorizzazioni Lake Formation su database e tabelle utilizzando istruzioni SQL modificate. Queste dichiarazioni sono simili alle dichiarazioni esistenti di Amazon Redshift. Per ulteriori informazioni, consulta GRANT and REVOKE nella Amazon Redshift Database Developer Guide.

Impostazione delle autorizzazioni per i formati di archiviazione a tabella aperta in Lake Formation

AWS Lake Formationsupporta la gestione delle autorizzazioni di accesso per Open Table Formats (OTF) come Apache Iceberg, Apache Hudi e Linux Foundation Delta Lake. In questo tutorial, imparerai come creare Iceberg, Hudi e Delta Lake utilizzando tabelle manifest symlinkAWS Glue, configurare autorizzazioni dettagliate AWS Glue Data Catalog utilizzando Lake Formation e interrogare i dati utilizzando Amazon Athena.

> **Note**
>
> AWSi servizi di analisi non supportano tutti i formati di tabelle transazionali. Per ulteriori informazioni, consulta Collaborazione con altri AWS servizi. Questo tutorial illustra manualmente la creazione di un nuovo database e di una tabella nel Data Catalog utilizzando solo i AWS Glue job.

Questo tutorial include un AWS CloudFormation modello per una configurazione rapida. Puoi rivederlo e personalizzarlo in base alle tue esigenze.

Argomenti

Destinatari principali

Questo tutorial è destinato agli amministratori IAM, agli amministratori di data lake e agli analisti aziendali. La tabella seguente elenca i ruoli utilizzati in questo tutorial per creare una tabella governata utilizzando Lake Formation.

Ruolo	Descrizione
Amministratore IAM	Un utente che può creare utenti e ruoli IAM e bucket Amazon S3. Ha la politica `AdministratorAccess` AWS gestita.
Amministratore del data lake	Un utente che può accedere al Data Catalog, creare database e concedere autorizzazioni Lake Formation ad altri utenti. Dispone di meno

Ruolo	Descrizione
	autorizzazioni IAM rispetto all'amministratore IAM, ma sufficienti per amministrare il data lake.
Business analyst	Un utente che può eseguire query sul data lake. Dispone delle autorizzazioni per eseguire query.

Prerequisiti

Prima di iniziare questo tutorial, devi disporre di un account Account AWS che ti consenta di accedere come utente con le autorizzazioni corrette. Per ulteriori informazioni, consultare Registrarsi per creare un Account AWS e Creazione di un utente amministratore.

Il tutorial presuppone che tu abbia familiarità con i ruoli e le politiche IAM. Per informazioni su IAM, consulta la IAM User Guide.

È necessario configurare le seguenti AWS risorse per completare questo tutorial:

- Utente amministratore di Data Lake
- Impostazioni del data lake Lake Formation
- Motore Amazon Athena versione 3

Per creare un amministratore di data lake

1. Accedi alla console di Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/ come utente amministratore. Creerai risorse nella regione Stati Uniti orientali (Virginia settentrionale) per questo tutorial.
2. Nella console di Lake Formation, nel riquadro di navigazione, in Autorizzazioni, scegli Ruoli e attività amministrativi.
3. Seleziona Scegli amministratori in Amministratori di Data lake.
4. Nella finestra pop-up, Gestisci gli amministratori del data lake, in Utenti e ruoli IAM, scegli Utente amministratore IAM.
5. Selezionare Salva.

Per abilitare le impostazioni del data lake

1. Aprire la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/. Nel riquadro di navigazione, in Catalogo dati, scegli Impostazioni. Deseleziona quanto segue:
   - Utilizza solo il controllo degli accessi IAM per i nuovi database.
   - Utilizza solo il controllo di accesso IAM per nuove tabelle in nuovi database.
2. In Impostazioni della versione per più account, scegli Versione 3 come versione per più account.
3. Selezionare Salva.

Per aggiornare il motore Amazon Athena alla versione 3

1. Apri la console Athena all'indirizzo https://console.aws.amazon.com/athena/.
2. Seleziona il gruppo di lavoro e seleziona il gruppo di lavoro principale.
3. Assicurati che il gruppo di lavoro abbia una versione minima di 3. In caso contrario, modifica il gruppo di lavoro, scegli Manual for Upgrade Query Engine e seleziona la versione 3.
4. Seleziona Salvataggio delle modifiche.

Fase 1: Fornisci le tue risorse

Questa sezione mostra come configurare le AWS risorse utilizzando un AWS CloudFormation modello.

Per creare le tue risorse utilizzando AWS CloudFormation un modello

1. Accedi alla AWS CloudFormation console all'indirizzo https://console.aws.amazon.com/cloudformation come amministratore IAM nella regione Stati Uniti orientali (Virginia settentrionale).
2. Scegli Launch Stack.
3. Scegli Avanti nella schermata Crea stack.
4. Inserisci un nome per lo stack.
5. Seleziona Avanti.
6. Nella pagina successiva, scegli Avanti.
7. Controlla i dettagli nella pagina finale e seleziona Riconosco che AWS CloudFormation potrebbe creare risorse IAM.

8. Scegli Crea.

   La creazione dello stack può richiedere fino a due minuti.

L'avvio dello stack di formazione cloud crea le seguenti risorse:

- lf-otf-datalake-123456789012 — Bucket Amazon S3 per archiviare dati

> **Note**
>
> L'ID dell'account aggiunto al nome del bucket Amazon S3 viene sostituito con l'ID dell'account.

- lf-otf-tutorial-123456789012 — Bucket Amazon S3 per archiviare i risultati delle query e gli script di lavoro AWS Glue
- AWS Gluelficebergdb — Database Iceberg
- lfhudidb — Banca dati AWS Glue Hudi
- lfdeltadb — Banca AWS Glue dati Delta
- native-iceberg-create — AWS Glue lavoro che crea una tabella Iceberg nel Data Catalog
- native-hudi-create — AWS Glue lavoro che crea una tabella Hudi nel Data Catalog
- native-delta-create — AWS Glue lavoro che crea una tabella Delta nel Data Catalog
- LF-OTF- GlueServiceRole — Ruolo IAM a cui si passa per AWS Glue eseguire i job. A questo ruolo sono allegate le politiche necessarie per accedere a risorse come Data Catalog, Amazon S3 bucket ecc.
- LF-OTF- RegisterRole — Ruolo IAM per registrare la posizione Amazon S3 con Lake Formation. Questo ruolo è `LF-Data-Lake-Storage-Policy` associato al ruolo.
- lf-consumer-analystuser — Utente IAM per interrogare i dati utilizzando Athena
- lf-consumer-analystuser-credentials — Password per l'utente dell'analista di dati memorizzata in AWS Secrets Manager

Una volta completata la creazione dello stack, vai alla scheda Output e annota i valori per:

- AthenaQueryResultLocation — Posizione Amazon S3 per l'output delle query Athena
- BusinessAnalystUserCredentials — Password per l'utente dell'analista di dati

Per recuperare il valore della password:

1. Scegli il `lf-consumer-analystuser-credentials` valore accedendo alla console Secrets Manager.
2. Nella sezione Secret value (Valore segreto) scegli Retrieve secret value (Recupera valore segreto).
3. Annota il valore segreto della password.

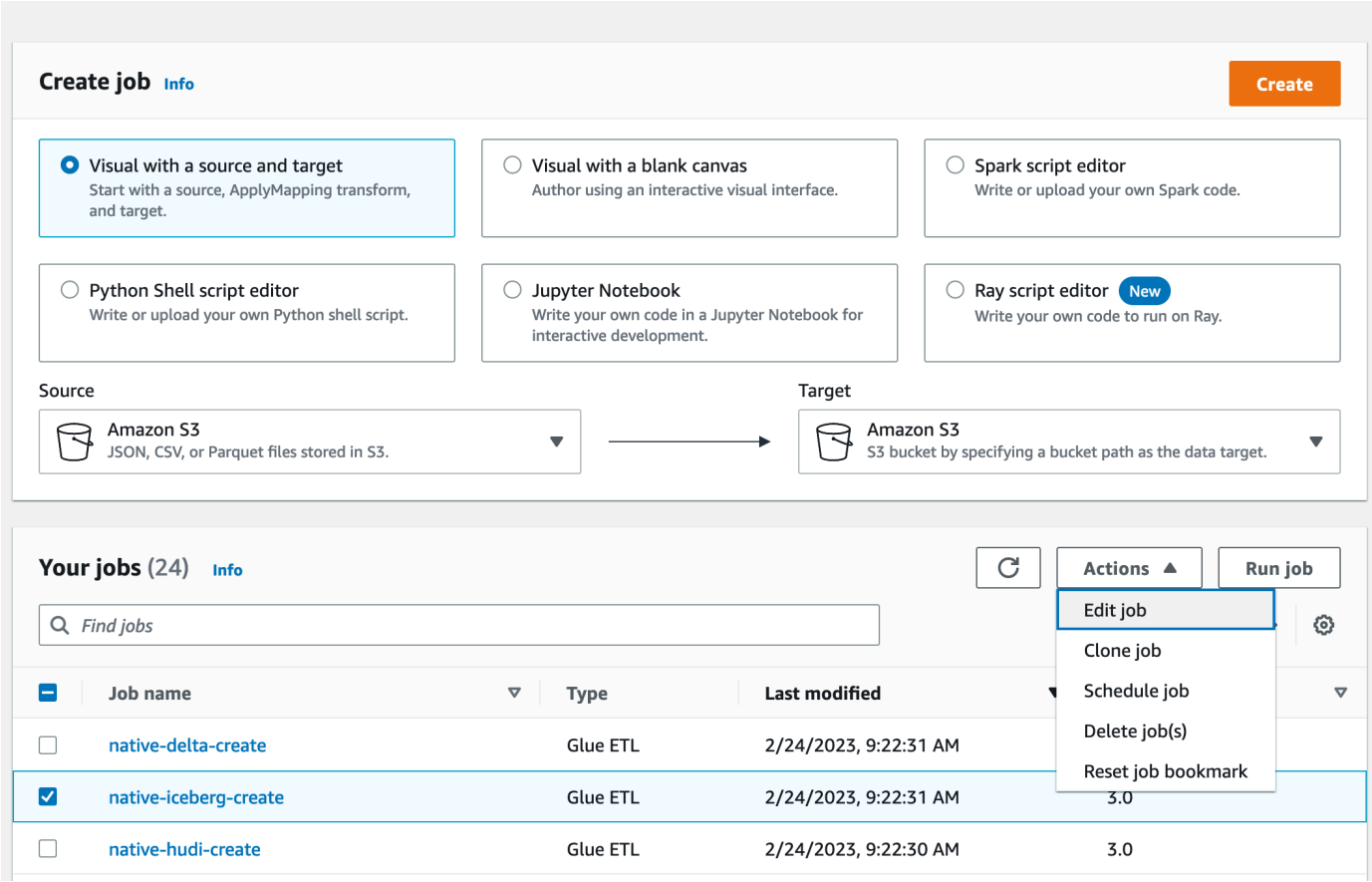
Passaggio 2: imposta le autorizzazioni per una tabella Iceberg

In questa sezione, imparerai come creare una tabella Iceberg inAWS Glue Data Catalog, impostare le autorizzazioni per i dati e interrogare i dati utilizzando Amazon Athena. AWS Lake Formation

Per creare una tabella Iceberg

In questo passaggio, eseguirai un AWS Glue job che crea una tabella transazionale Iceberg nel Data Catalog.

1. Apri la AWS Glue console all'indirizzo https://console.aws.amazon.com/glue/ nella regione Stati Uniti orientali (Virginia settentrionale) come utente amministratore del data lake.
2. Scegli i lavori dal riquadro di navigazione a sinistra.
3. Seleziona `native-iceberg-create`.



4. In Azioni, scegli Modifica lavoro.
5. In Dettagli Job, espandi Proprietà avanzate e seleziona la casella accanto a Usa AWS Glue Data Catalog come metastore Hive per aggiungere i metadati della tabella in. AWS Glue Data Catalog Questo specifica AWS Glue Data Catalog come metastore per le risorse del Data Catalog utilizzate nel job e consente di applicare successivamente le autorizzazioni di Lake Formation alle risorse del catalogo.
6. Selezionare Salva.
7. Scegli Esegui. È possibile visualizzare lo stato del job mentre è in esecuzione.

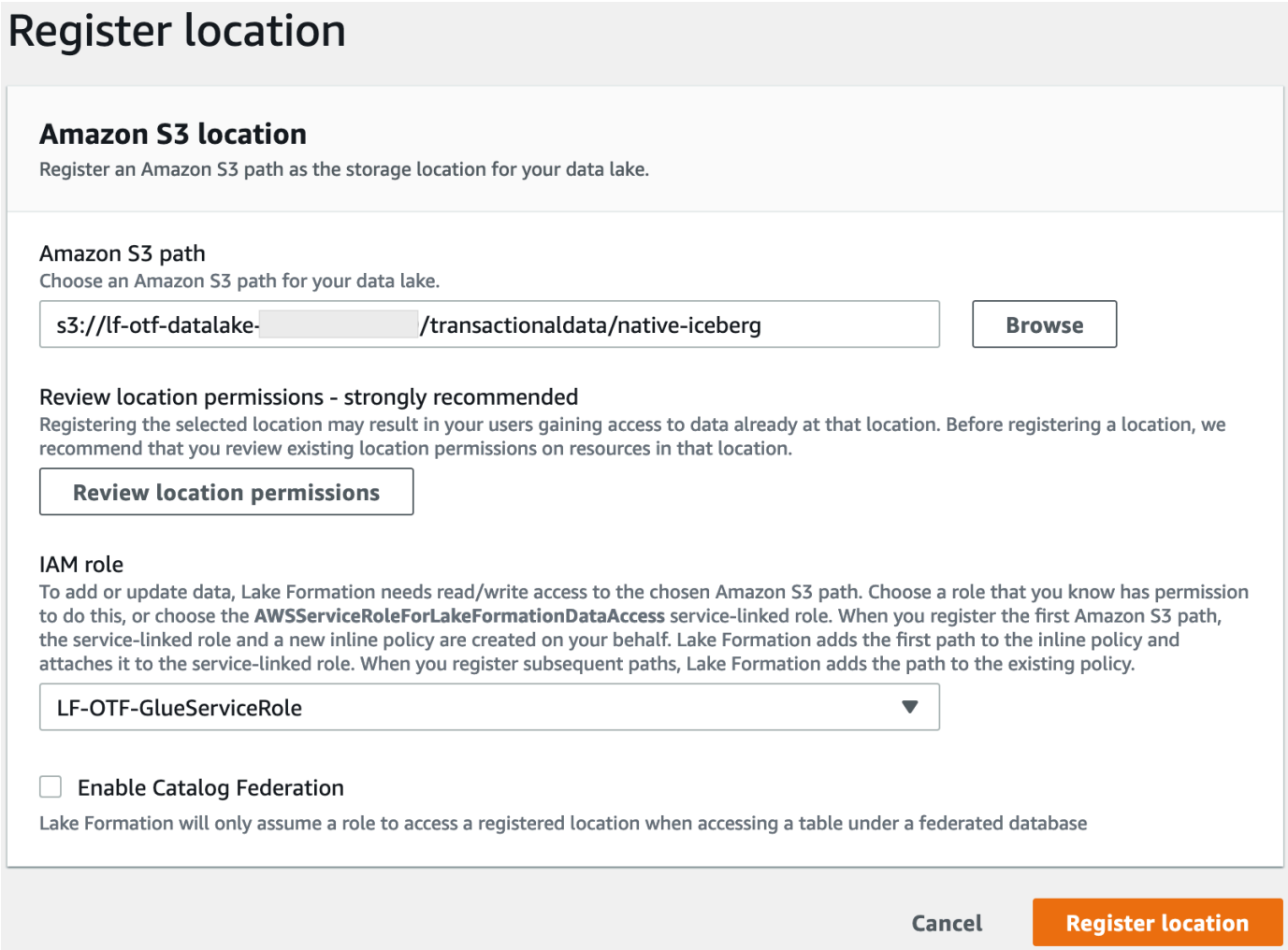
   Per ulteriori informazioni sui AWS Glue job, consulta Lavorare con i job sulla AWS Glue console nella AWS GlueDeveloper Guide.

   Questo lavoro crea una tabella Iceberg denominata `product` nel `lficebergdb` database. Verifica la tabella dei prodotti nella console Lake Formation.

Per registrare la posizione dei dati con Lake Formation

Successivamente, registra il percorso Amazon S3 come posizione del tuo data lake.

1. Apri la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/ come utente amministratore del data lake.
2. Nel pannello di navigazione, sotto Registra e immetti, scegli Posizione dei dati.
3. In alto a destra della console, scegli Registra posizione.
4. Nella pagina Registra posizione, inserisci quanto segue:
   - Percorso Amazon S3: scegli Sfoglia e seleziona. `lf-otf-datalake-123456789012` Fai clic sulla freccia destra (>) accanto alla posizione principale di Amazon S3 per accedere alla `s3/buckets/lf-otf-datalake-123456789012/transactionaldata/native-iceberg` posizione.
   - Ruolo IAM: scegli `LF-OTF-RegisterRole` come ruolo IAM.
   - Scegli la posizione di registrazione.

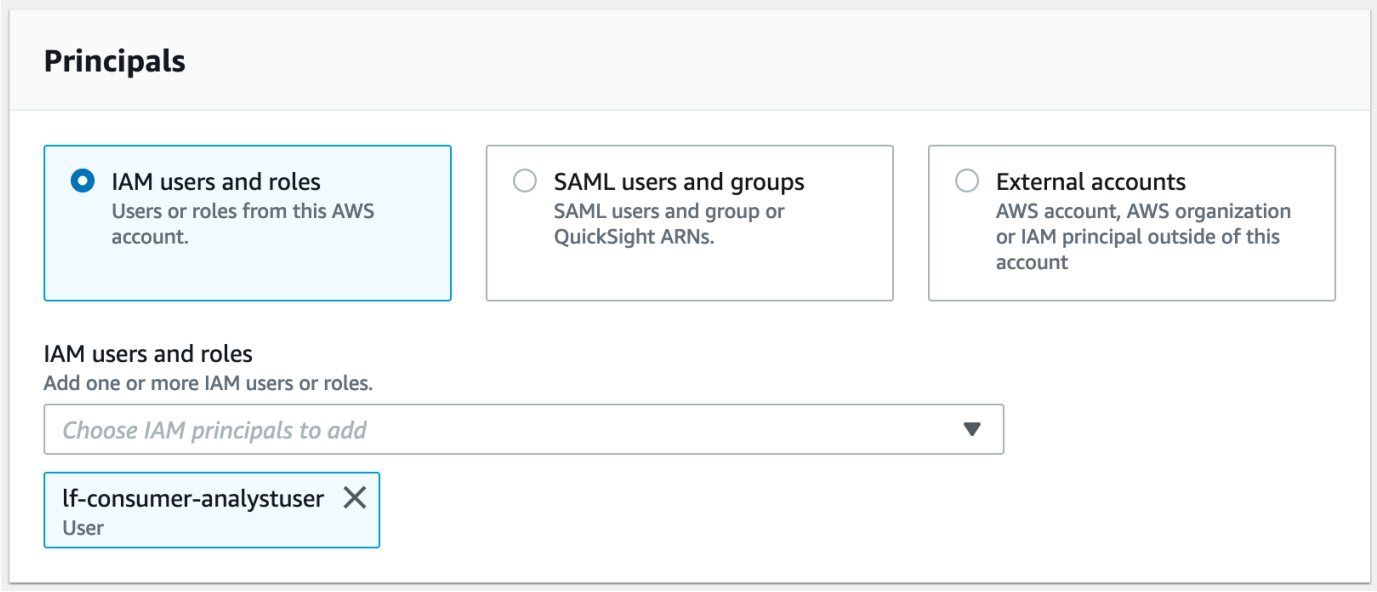


Per ulteriori informazioni sulla registrazione di una posizione dati con Lake Formation, vedereAggiungere una posizione Amazon S3 al tuo data lake.

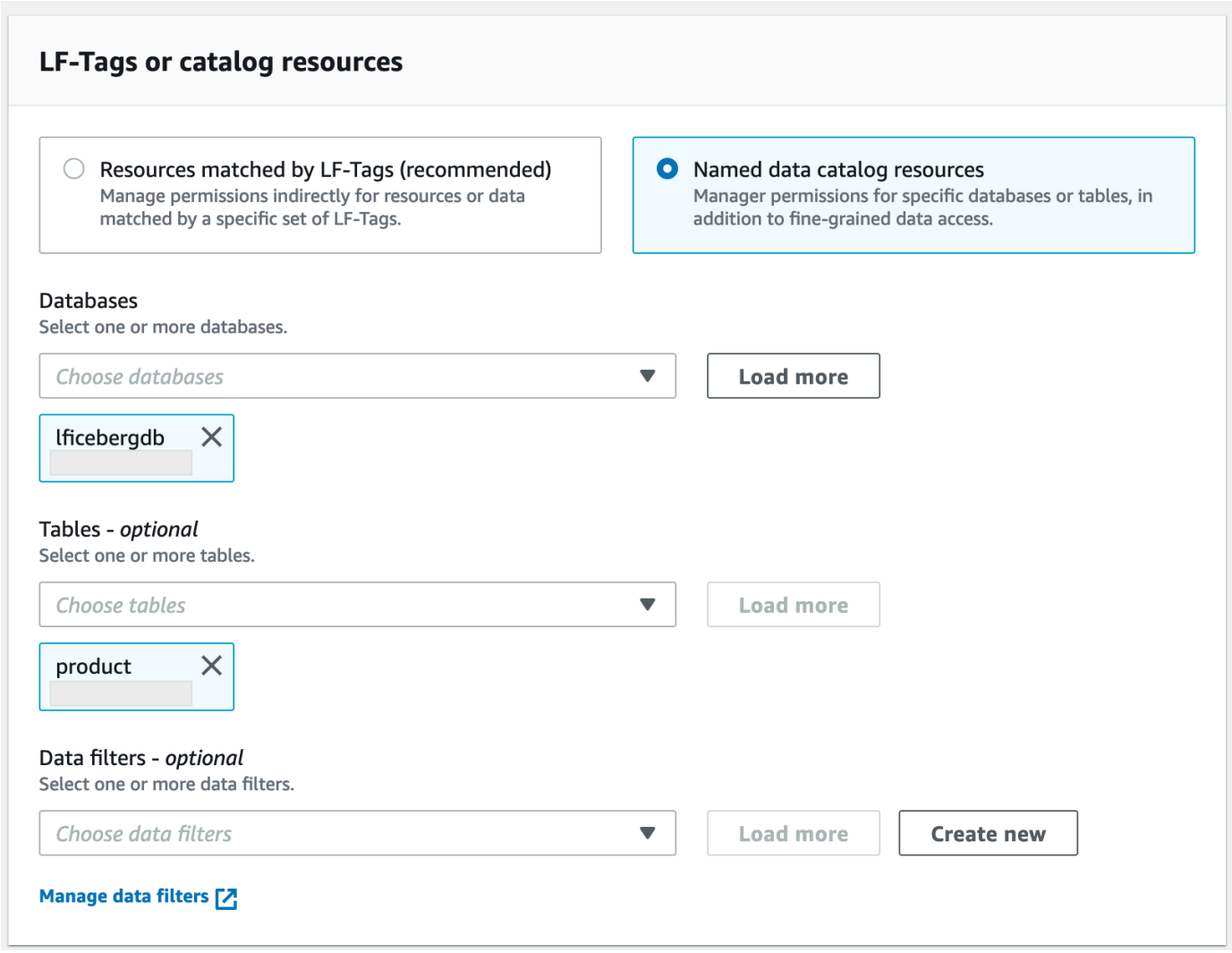
Per concedere i permessi di Lake Formation sulla tabella Iceberg

In questo passaggio, concederemo le autorizzazioni per il data lake all'utente business analyst.

1. In Autorizzazioni Data lake, scegli Concedi.
2. Nella schermata Concedi le autorizzazioni per i dati, scegli, utenti e ruoli IAM.
3. Scegli `lf-consumer-analystuser` dal menu a discesa.



4. Scegli la risorsa Named data catalog.
5. Per Database scegli`lficebergdb`.
6. Per Tabelle, scegli`product`.



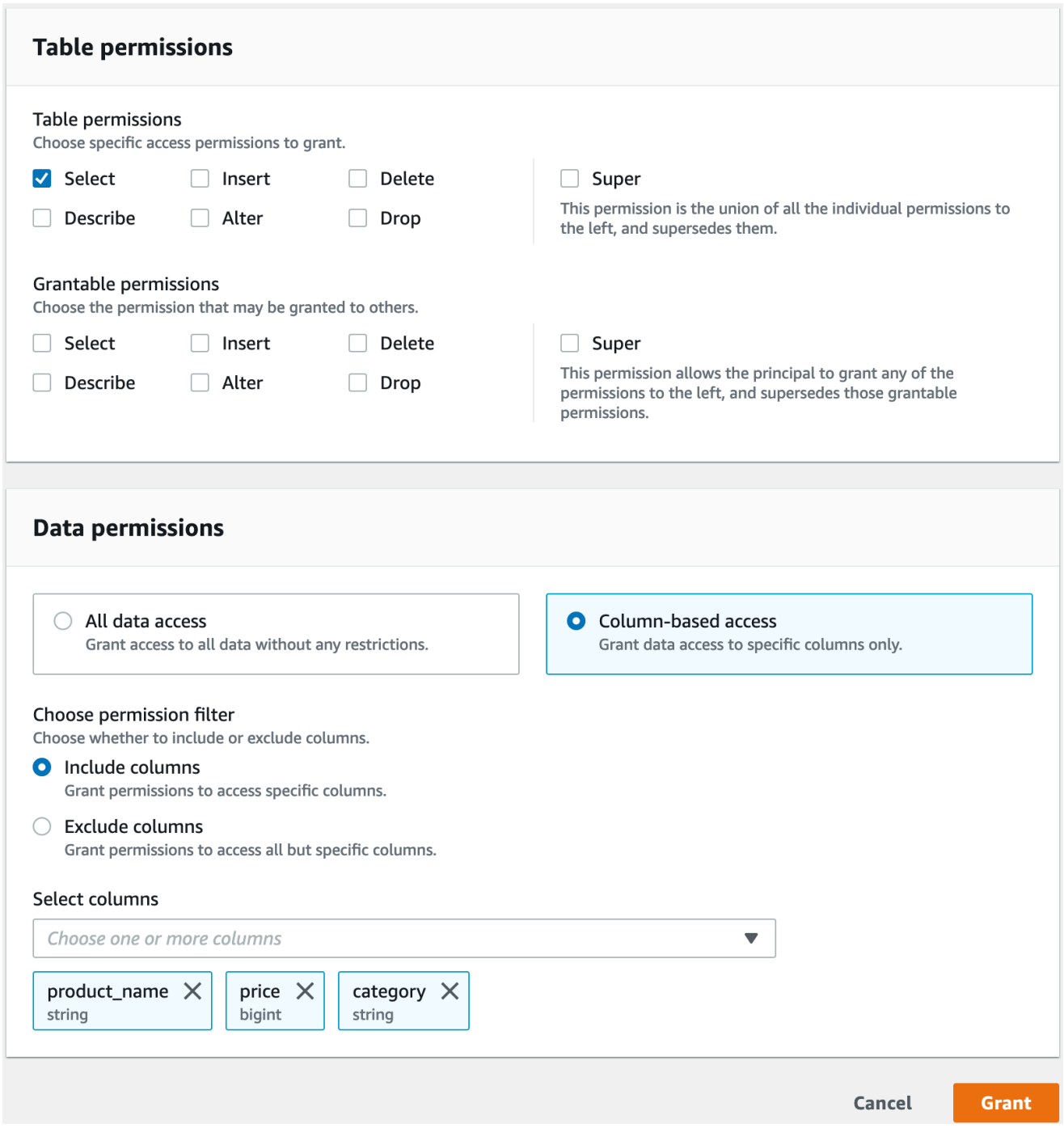
7. Successivamente, puoi concedere l'accesso basato su colonne specificando le colonne.

   a. In Autorizzazioni per la tabella, scegli Seleziona.

   b. In Autorizzazioni dati, scegli Accesso basato su colonne, scegli Includi colonne.

   c. Scegli`product_name`, e colonne`price`. `category`

   d. Scegli Concessione.



Per interrogare la tabella Iceberg usando Athena

Ora puoi iniziare a interrogare la tabella Iceberg che hai creato usando Athena. Se è la prima volta che esegui query in Athena, devi configurare una posizione dei risultati della query. Per ulteriori informazioni, vedere Specificazione della posizione dei risultati di una query.

1. Esci come utente amministratore del data lake e accedi come `lf-consumer-analystuser` nella regione Stati Uniti orientali (Virginia settentrionale) utilizzando la password indicata in precedenza nell'AWS CloudFormationoutput.
2. Aprire la console Athena all'indirizzo https://console.aws.amazon.com/athena/.
3. Scegli Impostazioni e seleziona Gestisci.
4. Nella casella Posizione dei risultati della query, inserisci il percorso del bucket che hai creato negli AWS CloudFormation output. Copia il valore di **`AthenaQueryResultLocation`** (s3://lf-otf-tutorial-123456789012/athena-results/) e scegli Salva.
5. Eseguite la seguente query per visualizzare in anteprima 10 record memorizzati nella tabella Iceberg:

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

   Per ulteriori informazioni sull'interrogazione delle tabelle Iceberg utilizzando Athena, consulta Interrogare le tabelle Iceberg nella Amazon Athena User Guide.

Passaggio 3: configura le autorizzazioni per una tabella Hudi

In questa sezione, imparerai come creare una tabella Hudi inAWS Glue Data Catalog, impostare le autorizzazioni per i dati e interrogare i dati utilizzando Amazon Athena. AWS Lake Formation

Per creare una tabella Hudi

In questo passaggio, eseguirai un AWS Glue job che crea una tabella transazionale Hudi nel Data Catalog.

1. Accedi alla AWS Glue console all'indirizzo https://console.aws.amazon.com/glue/ nella regione Stati Uniti orientali (Virginia settentrionale)

   come utente amministratore del data lake.
2. Scegli i lavori dal riquadro di navigazione a sinistra.
3. Seleziona `native-hudi-create`.
4. In Azioni, scegli Modifica lavoro.
5. In Dettagli Job, espandi Proprietà avanzate e seleziona la casella accanto a Usa AWS Glue Data Catalog come metastore Hive per aggiungere i metadati della tabella in. AWS Glue Data Catalog Questo specifica AWS Glue Data Catalog come metastore per le risorse del Data Catalog

utilizzate nel job e consente di applicare successivamente le autorizzazioni di Lake Formation alle risorse del catalogo.

6. Selezionare Salva.
7. Scegli Esegui. È possibile visualizzare lo stato del job mentre è in esecuzione.

   Per ulteriori informazioni sui AWS Glue job, consulta Lavorare con i job sulla AWS Glue console nella AWS GlueDeveloper Guide.

   Questo job crea una tabella Hudi (cow) nel database:lfhudidb. Verifica la `product` tabella nella console Lake Formation.

Per registrare la posizione dei dati con Lake Formation

Successivamente, registra un percorso Amazon S3 come posizione principale del tuo data lake.

1. Accedi alla console di Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/ come utente amministratore del data lake.
2. Nel pannello di navigazione, sotto Registra e inserisci, scegli Data location.
3. In alto a destra della console, scegli Registra posizione.
4. Nella pagina Registra posizione, inserisci quanto segue:

   - Percorso Amazon S3: scegli Sfoglia e seleziona. `lf-otf-datalake-123456789012` Fai clic sulla freccia destra (>) accanto alla posizione principale di Amazon S3 per accedere alla `s3/buckets/lf-otf-datalake-123456789012/transactionaldata/native-hudi` posizione.
   - Ruolo IAM: scegli `LF-OTF-RegisterRole` come ruolo IAM.
   - Scegli la posizione di registrazione.

Per concedere le autorizzazioni del data lake sulla tabella Hudi

In questo passaggio, concederemo le autorizzazioni del data lake all'utente business analyst.

1. In Autorizzazioni Data lake, scegli Concedi.
2. Nella schermata Concedi le autorizzazioni per i dati, scegli, utenti e ruoli IAM.
3. `lf-consumer-analystuser`dal menu a discesa.
4. Scegli la risorsa Named data catalog.

5. Per Database scegli`lfhudidb`.
6. Per Tabelle, scegli`product`.
7. Successivamente, puoi concedere l'accesso basato su colonne specificando le colonne.
   a. In Autorizzazioni per la tabella, scegli Seleziona.
   b. In Autorizzazioni dati, scegli Accesso basato su colonne, scegli Includi colonne.
   c. Scegli`product_name`, e colonne`price`. `category`
   d. Scegli Concessione.

Per interrogare la tabella Hudi usando Athena

Ora iniziate a interrogare la tabella Hudi che avete creato usando Athena. Se è la prima volta che esegui query in Athena, devi configurare una posizione dei risultati della query. Per ulteriori informazioni, vedere Specificazione della posizione dei risultati di una query.

1. Esci come utente amministratore del data lake e accedi come `lf-consumer-analystuser` nella regione Stati Uniti orientali (Virginia settentrionale) utilizzando la password indicata in precedenza nell'AWS CloudFormationoutput.
2. Aprire la console Athena all'indirizzo https://console.aws.amazon.com/athena/.
3. Scegli Impostazioni e seleziona Gestisci.
4. Nella casella Posizione dei risultati della query, inserisci il percorso del bucket che hai creato negli AWS CloudFormation output. Copia il valore di **`AthenaQueryResultLocation`** (s3://lf-otf-tutorial-123456789012/athena-results/) e salva.
5. Eseguite la seguente query per visualizzare in anteprima 10 record memorizzati nella tabella Hudi:

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

   Per ulteriori informazioni sull'interrogazione delle tabelle Hudi, consulta la sezione Interrogazione delle tabelle Hudi nella Guida per l'utente di Amazon Athena.

Passaggio 4: configura le autorizzazioni per una tabella Delta Lake

In questa sezione, imparerai come creare una tabella Delta Lake con il file manifest symlink nel fileAWS Glue Data Catalog, impostare le autorizzazioni per i dati AWS Lake Formation e interrogare i dati utilizzando Amazon Athena.

Per creare una tabella Delta Lake

In questo passaggio, eseguirai un AWS Glue processo che crea una tabella transazionale Delta Lake nel Data Catalog.

1. Accedi alla AWS Glue console all'indirizzo https://console.aws.amazon.com/glue/ nella regione Stati Uniti orientali (Virginia settentrionale)

   come utente amministratore del data lake.
2. Scegli i lavori dal riquadro di navigazione a sinistra.
3. Seleziona `native-delta-create`.
4. In Azioni, scegli Modifica lavoro.
5. In Dettagli Job, espandi Proprietà avanzate e seleziona la casella accanto a Usa AWS Glue Data Catalog come metastore Hive per aggiungere i metadati della tabella in. AWS Glue Data Catalog Questo specifica AWS Glue Data Catalog come metastore per le risorse del Data Catalog utilizzate nel job e consente di applicare successivamente le autorizzazioni di Lake Formation alle risorse del catalogo.
6. Selezionare Salva.
7. Scegli Esegui in Azioni.

   Questo lavoro crea una tabella Delta Lake denominata `product` nel `lfdeltadb` database. Verifica la `product` tabella nella console Lake Formation.

Per registrare la posizione dei dati con Lake Formation

Successivamente, registra il percorso Amazon S3 come posizione principale del tuo data lake.

1. Apri la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/, l'utente amministratore del data lake.
2. Nel pannello di navigazione, sotto Registra e inserisci, scegli Posizione dei dati.
3. In alto a destra della console, scegli Registra posizione.

4. Nella pagina Registra posizione, inserisci quanto segue:

   - Percorso Amazon S3: scegli Sfoglia e seleziona. `lf-otf-datalake-123456789012` Fai clic sulla freccia destra (**>**) accanto alla posizione principale di Amazon S3 per accedere alla `s3/buckets/lf-otf-datalake-123456789012/transactionaldata/native-delta` posizione.
   - Ruolo IAM: scegli `LF-OTF-RegisterRole` come ruolo IAM.
   - Scegli la posizione di registrazione.

Per concedere le autorizzazioni del data lake sulla tabella Delta Lake

In questo passaggio, concederemo le autorizzazioni del data lake all'utente business analyst.

1. In Autorizzazioni Data lake, scegli Concedi.
2. Nella schermata Concedi le autorizzazioni per i dati, scegli, utenti e ruoli IAM.
3. `lf-consumer-analystuser`dal menu a discesa.
4. Scegli la risorsa Named data catalog.
5. Per Database scegli`lfdeltadb`.
6. Per Tabelle, scegli`product`.
7. Successivamente, puoi concedere l'accesso basato su colonne specificando le colonne.

   a. In Autorizzazioni per la tabella, scegli Seleziona.

   b. In Autorizzazioni dati, scegli Accesso basato su colonne, scegli Includi colonne.

   c. Scegli`product_name`, e colonne`price`. `category`

   d. Scegli Concessione.

Per interrogare la tabella Delta Lake usando Athena

Ora iniziate a interrogare la tabella Delta Lake che avete creato usando Athena. Se è la prima volta che esegui query in Athena, devi configurare una posizione dei risultati della query. Per ulteriori informazioni, vedere Specificazione della posizione dei risultati di una query.

1. Esci come utente amministratore del data lake e accedi come `BusinessAnalystUser` nella regione Stati Uniti orientali (Virginia settentrionale) utilizzando la password indicata in precedenza nell'AWS CloudFormationoutput.

2. Aprire la console Athena all'indirizzo https://console.aws.amazon.com/athena/.
3. Scegli Impostazioni e seleziona Gestisci.
4. Nella casella Posizione dei risultati della query, inserisci il percorso del bucket che hai creato negli AWS CloudFormation output. Copia il valore di **`AthenaQueryResultLocation`** (s3://lf-otf-tutorial-123456789012/athena-results/) e salva.
5. Esegui la seguente query per visualizzare in anteprima 10 record memorizzati nella tabella Delta Lake:

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

Per ulteriori informazioni sull'interrogazione delle tabelle Delta Lake, consulta la sezione Interrogazione delle tabelle Delta Lake nella Guida per l'utente di Amazon Athena.

Fase 5: eliminazione delle risorse AWS

Per eliminare le risorse

Per evitare addebiti indesideratiAccount AWS, elimina le AWS risorse che hai utilizzato per questo tutorial.

1. Accedi alla AWS CloudFormation console all'indirizzo https://console.aws.amazon.com/cloudformation come amministratore IAM.
2. Elimina lo stack di formazione del cloud. Le tabelle che hai creato vengono eliminate automaticamente con lo stack.

Gestione di un data lake utilizzando il controllo degli accessi basato su tag Lake Formation

Migliaia di clienti stanno costruendo data lake su scala petabyte. AWS Molti di questi clienti lo utilizzano AWS Lake Formation per creare e condividere facilmente i propri data lake all'interno dell'organizzazione. Con l'aumento del numero di tabelle e utenti, i data steward e gli amministratori sono alla ricerca di modi per gestire le autorizzazioni sui data lake in modo semplice e su larga scala. Il controllo degli accessi basato su Lake Formation Tag (LF-TBAC) risolve questo problema consentendo ai data steward di creare tag LF (in base alla classificazione e all'ontologia dei dati) che possono quindi essere collegati alle risorse.

LF-TBAC è una strategia di autorizzazione che definisce le autorizzazioni in base agli attributi. In Lake Formation, questi attributi sono chiamati LF-tag. Puoi allegare i tag LF alle risorse del Data Catalog e ai principali di Lake Formation. Gli amministratori del Data Lake possono assegnare e revocare le autorizzazioni sulle risorse di Lake Formation utilizzando i tag LF. Per ulteriori informazioni su see,. Controllo degli accessi basato su tag Lake Formation

Questo tutorial dimostra come creare una politica di controllo degli accessi basata su tag Lake Formation utilizzando un set di dati AWS pubblico. Inoltre, mostra come interrogare tabelle, database e colonne a cui sono associate politiche di accesso basate su tag Lake Formation.

È possibile utilizzare LF-TBAC per i seguenti casi d'uso:

- Hai un gran numero di tabelle e principali a cui l'amministratore del data lake deve concedere l'accesso
- Vuoi classificare i tuoi dati in base a un'ontologia e concedere le autorizzazioni in base alla classificazione
- L'amministratore del data lake desidera assegnare le autorizzazioni in modo dinamico, in modo vagamente accoppiato

Di seguito sono riportati i passaggi di alto livello per la configurazione delle autorizzazioni utilizzando LF-TBAC:

1. Il data steward definisce l'ontologia dei tag con due tag LF: e. `Confidential` `Sensitive` Data with `Confidential=True` ha controlli di accesso più rigorosi. I dati `Sensitive=True` richiedono un'analisi specifica da parte dell'analista.
2. Il data steward assegna diversi livelli di autorizzazione al data engineer per creare tabelle con tag LF diversi.
3. L'ingegnere dei dati crea due database: e. `tag_database` `col_tag_database` Tutte le tabelle `tag_database` sono configurate con`Confidential=True`. Tutte le tabelle di `col_tag_database` sono configurate con`Confidential=False`. Alcune colonne della tabella `col_tag_database` sono contrassegnate con tag `Sensitive=True` per esigenze di analisi specifiche.
4. L'ingegnere dei dati concede il permesso di lettura all'analista per le tabelle con condizioni di espressione specifiche `Confidential=True` e`Confidential=False`,. `Sensitive=True`
5. Con questa configurazione, l'analista di dati può concentrarsi sull'esecuzione dell'analisi con i dati giusti.

Argomenti

Destinatari principali

Questo tutorial è destinato a amministratori di dati, ingegneri di dati e analisti di dati. Per quanto riguarda la gestione AWS Glue Data Catalog e l'amministrazione delle autorizzazioni in Lake Formation, gli amministratori dei dati all'interno degli account di produzione hanno la proprietà funzionale in base alle funzioni che supportano e possono concedere l'accesso a vari consumatori, organizzazioni esterne e account.

La tabella seguente elenca i ruoli utilizzati in questo tutorial:

Ruolo	Descrizione
Data steward (amministratore)	L'`lf-data-steward` utente dispone dei seguenti accessi: • Accesso in lettura a tutte le risorse del Data Catalog • Può creare tag LF e associarsi al ruolo di ingegnere dei dati per concedere l'autoriz zazione ad altri responsabili
Ingegnere dei dati	`lf-data-engineer` l'utente ha i seguenti accessi:

Ruolo	Descrizione
	• Accesso completo in lettura, scrittura e aggiornamento a tutte le risorse del Data Catalog • Autorizzazioni per la localizzazione dei dati nel data lake • Può associare tag LF e associarsi al Data Catalog • Può allegare tag LF alle risorse, il che fornisce l'accesso ai principali in base a qualsiasi politica creata dai data steward
Analista dei dati	L'`lf-data-analyst` utente ha il seguente accesso: • Accesso granulare alle risorse condivise dalle politiche di accesso basate su tag di Lake Formation

Prerequisiti

Prima di iniziare questo tutorial, è necessario disporre di un file Account AWS da utilizzare per accedere come utente amministrativo con le autorizzazioni corrette. Per ulteriori informazioni, consulta Completa i processi di configurazione iniziali AWS.

Il tutorial presuppone che tu abbia familiarità con IAM. Per informazioni su IAM, consulta la IAM User Guide.

Fase 1: Fornisci le tue risorse

Questo tutorial include un AWS CloudFormation modello per una configurazione rapida. Puoi rivederlo e personalizzarlo in base alle tue esigenze. Il modello crea tre ruoli diversi (elencati inDestinatari principali) per eseguire questo esercizio e copia il nyc-taxi-data set di dati nel bucket Amazon S3 locale.

- Un bucket Amazon S3

- Le impostazioni appropriate di Lake Formation
- Le risorse Amazon EC2 appropriate
- Tre ruoli IAM con credenziali

Crea le tue risorse

1. Accedi alla AWS CloudFormation console all'indirizzo https://console.aws.amazon.com/cloudformation nella regione Stati Uniti orientali (Virginia settentrionale).
2. Scegli Launch Stack.
3. Seleziona Avanti.
4. Nella sezione Configurazione utente, inserisci la password per tre ruoli:`DataStewardUserPassword`, `DataEngineerUserPassword` e`DataAnalystUserPassword`.
5. Controlla i dettagli nella pagina finale e seleziona Riconosco che AWS CloudFormation potrebbe creare risorse IAM.
6. Scegli Crea.

   La creazione dello stack può richiedere fino a cinque minuti.

> **Note**
>
> Dopo aver completato il tutorial, potresti voler eliminare lo stack per evitare di continuare AWS CloudFormation a incorrere in addebiti. Verifica che le risorse siano state eliminate correttamente nello stato dell'evento per lo stack.

Fase 2: Registrare la posizione dei dati, creare un'ontologia LF-tag e concedere le autorizzazioni

In questo passaggio, l'utente Data Steward definisce l'ontologia dei tag con due LF-Tag: `Confidential` e `Sensitive` fornisce a specifici principi IAM la possibilità di allegare tag LF appena creati alle risorse.

Registra una posizione dei dati e definisci l'ontologia LF-Tag

1. Esegui il primo passaggio come utente amministratore dei dati (`lf-data-steward`) per verificare i dati in Amazon S3 e nel Data Catalog in Lake Formation.

   a. Accedi alla console di Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/ `lf-data-steward` con la password utilizzata durante la distribuzione dello AWS CloudFormation stack.

   b. Nel riquadro di navigazione, in Autorizzazioni, scegli Ruoli e attività amministrative.

   c. Scegli Aggiungi nella sezione Amministratori di Data Lake.

   d. Nella pagina Aggiungi amministratore, per gli utenti e i ruoli IAM, scegli l'utente`lf-data-steward`.

   e. Scegli Salva per aggiungerlo `lf-data-steward` come amministratore di Lake Formation.

2. Successivamente, aggiorna le impostazioni del Data Catalog per utilizzare l'autorizzazione Lake Formation per controllare le risorse del catalogo anziché il controllo degli accessi basato su IAM.

   a. Nel pannello di navigazione, in Amministrazione, scegli Impostazioni Data Catalog.

   b. Deseleziona Usa solo il controllo di accesso IAM per i nuovi database.

   c. Deseleziona Usa solo il controllo di accesso IAM per nuove tabelle in nuovi database.

   d. Fai clic su Save (Salva).

3. Quindi, registra la posizione dei dati per il data lake.

   a. Nel riquadro di navigazione, in Amministrazione, scegli Posizioni dei data lake.

   b. Scegli Registra posizione.

   c. Nella pagina Registra posizione, per il percorso Amazon S3, inserisci. `s3://lf-tagbased-demo-`*`Account-ID`*

   d. Per il ruolo IAM, lascia il valore predefinito `AWSServiceRoleForLakeFormationDataAccess` così com'è.

   e. Scegli Lake Formation come modalità di autorizzazione.

   f. Scegli Registra posizione.

4. Quindi, crea l'ontologia definendo un tag LF.

   a. In Autorizzazioni nel riquadro di navigazione, scegli LF-tags e permessi. .

   b. Scegli Aggiungi tag LF.

   c. In Chiave, inserire `Confidential`.

d. Per Valori, aggiungi `True` e. `False`

e. Scegli Aggiungi tag LF.

f. Ripeti i passaggi per creare il tag LF `Sensitive` con il valore. `True`

Avete creato tutti i tag LF necessari per questo esercizio.

Concedi le autorizzazioni agli utenti IAM

1. Successivamente, offri a specifici principi IAM la possibilità di allegare tag LF appena creati alle risorse.

   a. In Autorizzazioni nel pannello di navigazione, scegli LF-tags e permessi.

   b. Nella sezione Autorizzazioni LF-Tag, scegli Concedi autorizzazioni.

   c. Per Tipo di autorizzazione, scegli le autorizzazioni per la coppia chiave-valore LF-Tag.

   d. Seleziona utenti e ruoli IAM.

   e. Per gli utenti e i ruoli IAM, cerca e scegli il `lf-data-engineer` ruolo.

   f. Nella sezione LF-Tags, aggiungi la chiave `Confidential` con valori `True` e `False` e `key Sensitive` with value. `True`

   g. In Autorizzazioni, seleziona Descrivi e associa per le autorizzazioni e le autorizzazioni concedibili.

   h. Scegli Concessione.

2. Successivamente, concedi le autorizzazioni `lf-data-engineer` per creare database nel nostro catalogo dati e nel bucket Amazon S3 sottostante creato da. AWS CloudFormation

   a. In Amministrazione nel riquadro di navigazione, scegli Ruoli e attività amministrative.

   b. Nella sezione Creatori di database, scegli Concedi.

   c. Per gli utenti e i ruoli IAM, scegli il `lf-data-engineer` ruolo.

   d. Per le autorizzazioni del catalogo, seleziona Crea database.

   e. Scegli Concessione.

3. Successivamente, concedi all'utente le autorizzazioni sul (`s3://lf-tagbased-demo-`*`Account-ID`*`)` bucket Amazon S3. `lf-data-engineer`

   a. Nel riquadro di navigazione, in Autorizzazioni, scegli Posizioni dei dati.

   b. Scegli Concessione.

c. Seleziona Il mio account.

d. Per gli utenti e i ruoli IAM, scegli il `lf-data-engineer` ruolo.

e. Per le posizioni di archiviazione, inserisci il bucket Amazon S3 creato dal modello. AWS CloudFormation (`s3://lf-tagbased-demo-`*`Account-ID`*`)`

f. Scegli Concessione.

4. Successivamente, concedi autorizzazioni **`lf-data-engineer`** concedibili sulle risorse associate all'espressione LF-Tag. `Confidential=True`

   a. Nel riquadro di navigazione, sotto Autorizzazioni, scegli Autorizzazioni Data lake.

   b. Scegli Concessione.

   c. Seleziona utenti e ruoli IAM.

   d. Scegli il ruolo`lf-data-engineer`.

   e. Nella sezione LF-tags o risorse del catalogo, seleziona Risorse abbinate ai tag LF.

   f. Scegliete Aggiungi coppia chiave-valore LF-Tag.

   g. Aggiungi la chiave con i valori`Confidential`. `True`

   h. Nella sezione Autorizzazioni del database, seleziona Descrivi per le autorizzazioni del database e le autorizzazioni concedibili.

   i. Nella sezione Autorizzazioni per la tabella, seleziona Descrivi, Seleziona e Alter sia per le autorizzazioni Table che per le autorizzazioni Grantable.

   j. Scegli Concessione.

5. Successivamente, concedi autorizzazioni `lf-data-engineer` concedibili sulle risorse associate all'espressione LF-Tag. `Confidential=False`

   a. Nel riquadro di navigazione, sotto Autorizzazioni, scegli Autorizzazioni Data lake.

   b. Scegli Concessione.

   c. Seleziona utenti e ruoli IAM.

   d. Scegli il ruolo`lf-data-engineer`.

   e. Seleziona Risorse abbinate ai tag LF.

   f. Scegli Aggiungi tag LF.

   g. Aggiungi la chiave `Confidential` con il valore. `False`

   h. Nella sezione Autorizzazioni del database, seleziona Descrivi per le autorizzazioni del database e le autorizzazioni concedibili.

i. Nella sezione Autorizzazioni per tabelle e colonne, non selezionare nulla.

j. Scegli Concessione.

6. Successivamente, concediamo autorizzazioni `lf-data-engineer` concedibili sulle risorse associate alle coppie chiave-valore del tag LF e. `Confidential=False` `Sensitive=True`

a. Nel pannello di navigazione, in Autorizzazioni, scegli Autorizzazioni dati.

b. Scegli Concessione.

c. Seleziona utenti e ruoli IAM.

d. Scegli il ruolo`lf-data-engineer`.

e. Nella sezione LF-tags o nella sezione delle risorse del catalogo, seleziona Risorse abbinate ai tag LF.

f. Scegli Aggiungi tag LF.

g. Aggiungi la chiave `Confidential` con il valore. `False`

h. Scegli Aggiungi coppia chiave-valore del tag LF.

i. Aggiungi la chiave con il valore`Sensitive`. `True`

j. Nella sezione Autorizzazioni del database, seleziona Descrivi per le autorizzazioni del database e le autorizzazioni concedibili.

k. Nella sezione Autorizzazioni per la tabella, seleziona Descrivi, Seleziona e Alter sia per le autorizzazioni Table che per le autorizzazioni Grantable.

l. Scegli Concessione.

Fase 3: Creare database Lake Formation

In questo passaggio, create due database e allegate i tag LF ai database e a colonne specifiche a scopo di test.

Crea i tuoi database e le tue tabelle per l'accesso a livello di database

1. Innanzitutto, crea il database`tag_database`, la tabella e allega i tag `source_data` LF appropriati.

a. Nella console Lake Formation (https://console.aws.amazon.com/lakeformation/), in Data Catalog, scegli Databases.

b. Scegliere Crea database.

c. Per Nome, immetti `tag_database`.

d. Per Posizione, inserisci la posizione Amazon S3 creata dal AWS CloudFormation modello. (`s3://lf-tagbased-demo-`*`Account-ID`*`/tag_database/`)

e. Deseleziona Usa solo il controllo di accesso IAM per le nuove tabelle in questo database.

f. Scegliere Crea database.

2. Quindi, crea una nuova tabella all'interno `tag_database` di.

a. Nella pagina Database, seleziona il database`tag_database`.

b. Scegliete Visualizza tabelle e fate clic su Crea tabella.

c. Per Nome, immetti `source_data`.

d. Per Database (Database), scegli il database `tag_database`.

e. Per Formato tabella, scegli AWS GlueTabella standard.

f. Se i dati si trovano in, seleziona Percorso specificato nel mio account.

g. In Includi percorso, inserisci il percorso `tag_database` creato dal AWS CloudFormation modello(`s3://lf-tagbased-demo`*`Account-ID`*`/tag_database/`).

h. Per Formato dati, seleziona CSV.

i. In Upload schema, inserisci la seguente matrice JSON di struttura di colonne per creare uno schema:

```
[
              {
                   "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. Scegli Carica. Dopo aver caricato lo schema, lo schema della tabella dovrebbe apparire come nella schermata seguente:

#	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. Seleziona Invia.

3. Successivamente, allega i tag LF a livello di database.

   a. Nella pagina Database, trova e seleziona. `tag_database`
   b. Nel menu Azioni, scegliete Modifica tag LF.
   c. Scegliete Assegna nuovo LF-tag.
   d. Per le chiavi assegnate, scegliete il tag `Confidential` LF che avete creato in precedenza.
   e. Per Valori, scegliete. `True`
   f. Selezionare Salva.

   Questo completa l'assegnazione del tag LF al database tag_database.

Crea il tuo database e la tua tabella per l'accesso a livello di colonna

Ripetete i seguenti passaggi per creare il database `col_tag_database` e la tabella e allegare `source_data_col_lvl` i tag LF a livello di colonna.

1. Nella pagina Database, scegli Crea database.
2. Per Nome, immetti `col_tag_database`.
3. Per Posizione, inserisci la posizione Amazon S3 creata dal AWS CloudFormation modello. (`s3://lf-tagbased-demo-`*`Account-ID`*`/col_tag_database/`)
4. Deseleziona Usa solo il controllo di accesso IAM per le nuove tabelle in questo database.
5. Scegliere Crea database.
6. Nella pagina Database, seleziona il tuo nuovo database(`col_tag_database`).
7. Scegli Visualizza tabelle e fai clic su Crea tabella.
8. Per Nome, immetti `source_data_col_lvl`.
9. Per Database, scegli il tuo nuovo database(`col_tag_database`).
10. Per Formato tabella, scegli AWS GlueTabella standard.
11. Se i dati si trovano in, seleziona Percorso specificato nel mio account.
12. Inserisci il percorso Amazon S3 per. `col_tag_database` (`s3://lf-tagbased-demo-`*`Account-ID`*`/col_tag_database/`)
13. Per Formato dei dati, selezionaCSV.
14. In`Upload schema`, inserisci il seguente schema 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. Scegli `Upload`. Dopo aver caricato lo schema, lo schema della tabella dovrebbe apparire come nella schermata seguente.

#	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. Scegli Invia per completare la creazione della tabella.
17. Ora, associa il `Sensitive=True` tag LF alle colonne `vendorid` e. `fare_amount`

    a. Nella pagina Tabelle, seleziona la tabella che hai creato. `(source_data_col_lvl)`

    b. Nel menu Azioni, scegli Schema.

    c. Selezionate la colonna `vendorid` e scegliete Modifica tag LF.

    d. Per Chiavi assegnate, scegliete Sensibile.

    e. Per Valori, scegli True.

    f. Selezionare Salva.
18. Quindi, associa il `Confidential=False` tag LF a. `col_tag_database` Questo è necessario per poter `lf-data-analyst` descrivere il database `col_tag_database` quando si effettua l'accesso da. Amazon Athena

    a. Nella pagina Database, trova e selezionacol_tag_database.

    b. Nel menu Azioni, scegliete Modifica tag LF.

    c. Scegliete Assegna nuovo LF-tag.

    d. Per Tasti assegnati, scegli il tag `Confidential` LF che hai creato in precedenza.

    e. Per Valori, scegliete. `False`

    f. Selezionare Salva.

Fase 4: Concedere le autorizzazioni relative alla tabella

Concedi le autorizzazioni agli analisti di dati per l'utilizzo dei database `tag_database` e della tabella `col_tag_database` utilizzando i tag LF e. `Confidential` `Sensitive`

1. Segui questi passaggi per concedere all'`lf-data-analyst`utente le autorizzazioni sugli oggetti associati al tag LF `Confidential=True` (Database:TAG_database) per avere il database e l'autorizzazione sulle tabelle. `Describe` `Select`

    a. Accedi alla console di Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/ com`lf-data-engineer`.

    b. In Autorizzazioni, seleziona Autorizzazioni Data lake.

    c. Scegli Concessione.

    d. In Principali, seleziona Utenti e ruoli IAM.

e. Per gli utenti e i ruoli IAM, scegli`lf-data-analyst`.

f. In LF-tags o catalog resources, seleziona Risorse abbinate a LF-Tags.

g. Scegli Aggiungi tag LF.

h. Per Key, scegli. `Confidential`

i. Per Valori, scegli`True`.

j. Per le autorizzazioni del database, selezionare`Describe`.

k. Per le autorizzazioni relative alle tabelle, scegli Seleziona e descrivi.

l. Scegli Concessione.

2. Quindi, ripeti i passaggi per concedere le autorizzazioni agli analisti di dati per l'espressione LF-tag for. `Confidential=False` Questo tag LF viene utilizzato per descrivere la tabella `col_tag_database` e `source_data_col_lvl` quando si effettua l'accesso da Amazon `lf-data-analyst` Athena.

   a. Accedi alla console di Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/ come`lf-data-engineer`.

   b. Nella pagina Database, seleziona il database`col_tag_database`.

   c. Scegli Azione e Concedi.

   d. In Principi, seleziona Utenti e ruoli IAM.

   e. Per gli utenti e i ruoli IAM, scegli`lf-data-analyst`.

   f. Seleziona Risorse abbinate ai tag LF.

   g. Scegli Aggiungi tag LF.

   h. Per Key, scegli. `Confidential`

   i. Per Valori, scegli`False`.

   j. Per le autorizzazioni del database, seleziona`Describe`.

   k. Per le autorizzazioni relative alle tabelle, non selezionate nulla.

   l. Scegli Concessione.

3. Quindi, ripeti i passaggi per concedere le autorizzazioni agli analisti di dati per l'espressione LF-tag per e. `Confidential=False` `Sensitive=True` Questo tag LF viene utilizzato per descrivere la tabella `col_tag_database` e `source_data_col_lvl` (a livello di colonna) quando si effettua l'accesso da Amazon Athena. `lf-data-analyst`

   a. Accedi alla console di Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/ come`lf-data-engineer`.

b. Nella pagina Database, seleziona il databasecol_tag_database.
c. Scegli Azione e Concedi.
d. In Principi, seleziona Utenti e ruoli IAM.
e. Per gli utenti e i ruoli IAM, sceglilf-data-analyst.
f. Seleziona Risorse abbinate ai tag LF.
g. Scegli Aggiungi tag LF.
h. Per Key, scegli. Confidential
i. Per Valori, scegliFalse.
j. Scegli Aggiungi tag LF.
k. Per Key, scegli. Sensitive
l. Per Valori, scegliTrue.
m. Per le autorizzazioni del database, selezionaDescribe.
n. Per le autorizzazioni relative alle tabelle, seleziona Select e. Describe
o. Scegli Concessione.

Passaggio 5: esegui una query in Amazon Athena per verificare le autorizzazioni

Per questo passaggio, usa Amazon Athena per eseguire SELECT query sulle due tabelle. (source_data and source_data_col_lvl) Usa il percorso Amazon S3 come posizione dei risultati della query. (s3://lf-tagbased-demo-*Account-ID*/athena-results/)

1. Accedi alla console Athena all'indirizzo https://console.aws.amazon.com/athena/ come. lf-data-analyst
2. Nell'editor di query Athena, scegli tag_database nel pannello a sinistra.
3. Scegliete l'icona delle opzioni di menu aggiuntive (tre punti verticali) accanto source_data e scegliete la tabella di anteprima.
4. Scegli Esegui query.

   L'esecuzione della query dovrebbe richiedere alcuni minuti. La query visualizza tutte le colonne dell'output perché il tag LF è associato a livello di database e la source_data tabella lo eredita automaticamente LF-tag dal database. tag_database
5. Esegui un'altra query utilizzando e. col_tag_database source_data_col_lvl

La seconda query restituisce le due colonne contrassegnate come `Non-Confidential` e`Sensitive`.

6. Puoi anche controllare il comportamento della politica di accesso basata sui tag di Lake Formation nelle colonne per le quali non hai concessioni politiche. Quando viene selezionata una colonna senza tag dalla tabella, `source_data_col_lvl` Athena restituisce un errore. Ad esempio, puoi eseguire la seguente query per scegliere colonne senza tag: `geolocationid`

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

Fase 6: Pulizia delle risorse AWS

Per evitare addebiti indesideratiAccount AWS, puoi eliminare le AWS risorse che hai utilizzato per questo tutorial.

1. Accedi alla console di Lake Formation come `lf-data-engineer` ed elimina i database `tag_database` e`col_tag_database`.
2. Successivamente, accedi come `lf-data-steward` e ripulisci tutte le autorizzazioni LF-Tag, le autorizzazioni per i dati e le autorizzazioni per la localizzazione dei dati concesse in precedenza e che sono state concesse e. `lf-data-engineer` `lf-data-analyst.`
3. Accedi alla console Amazon S3 come proprietario dell'account utilizzando le credenziali IAM che hai usato per distribuire lo stack. AWS CloudFormation
4. Elimina i seguenti bucket:
   - lf-tagbased-demo-accesslogs*- acct-id*
   - lf-tagbased-demo- *acct-id*
5. Accedi alla AWS CloudFormation console all'indirizzo https://console.aws.amazon.com/cloudformation ed elimina lo stack che hai creato. Attendi che lo stato dello stack cambi a. `DELETE_COMPLETE`

Protezione dei data lake con il controllo degli accessi a livello di riga

AWS Lake Formationle autorizzazioni a livello di riga consentono di fornire l'accesso a righe specifiche di una tabella in base alla conformità dei dati e alle politiche di governance. Se disponi di tabelle di grandi dimensioni che contengono miliardi di record, hai bisogno di un modo per consentire

a diversi utenti e team di accedere solo ai dati che sono autorizzati a vedere. Il controllo degli accessi a livello di riga è un modo semplice e performante per proteggere i dati, offrendo al contempo agli utenti l'accesso ai dati di cui hanno bisogno per svolgere il proprio lavoro. Lake Formation fornisce audit centralizzati e report di conformità identificando quali responsabili hanno avuto accesso a quali dati, quando e attraverso quali servizi.

In questo tutorial, imparerai come funzionano i controlli di accesso a livello di riga in Lake Formation e come configurarli.

Questo tutorial include un AWS CloudFormation modello per configurare rapidamente le risorse richieste. Puoi rivederlo e personalizzarlo in base alle tue esigenze.

Argomenti

Destinatari principali

Questo tutorial è destinato a amministratori di dati, ingegneri dei dati e analisti di dati. La tabella seguente elenca i ruoli e le responsabilità di un proprietario e di un consumatore di dati.

Ruolo	Descrizione
Amministratore IAM	Un utente che può creare utenti e ruoli e bucket Amazon Simple Storage Service (Amazon S3). Dispone della politica `AdministratorAccess` AWS gestita.
Amministratore del data lake	Un utente responsabile della configurazione del data lake, della creazione di filtri di dati e della

Ruolo	Descrizione
	concessione delle autorizzazioni agli analisti di dati.
Analista dei dati	Un utente che può eseguire query sul data lake. Gli analisti di dati residenti in diversi paesi (per il nostro caso d'uso, Stati Uniti e Giappone) possono analizzare solo le recensioni dei prodotti per i clienti residenti nel proprio paese e, per motivi di conformità, non dovrebbero essere in grado di vedere i dati dei clienti che si trovano in altri paesi.

Prerequisiti

Prima di iniziare questo tutorial, è necessario disporre di un file Account AWS da utilizzare per accedere come utente amministrativo con le autorizzazioni corrette. Per ulteriori informazioni, consulta Completa i processi di configurazione iniziali AWS.

Il tutorial presuppone che tu abbia familiarità con IAM. Per informazioni su IAM, consulta la IAM User Guide.

Modifica le impostazioni di Lake Formation

> **Important**
>
> Prima di avviare il AWS CloudFormation modello, disabilita l'opzione Usa solo il controllo di accesso IAM per nuovi database/tabelle in Lake Formation seguendo i passaggi seguenti:

1. Accedi alla console di Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/ nella regione Stati Uniti orientali (Virginia settentrionale) o Stati Uniti occidentali (Oregon).
2. In Data Catalog, scegli Impostazioni.
3. Deseleziona Usa solo il controllo di accesso IAM per i nuovi database e Usa solo il controllo di accesso IAM per le nuove tabelle nei nuovi database.

4. Seleziona Salva.

Fase 1: Fornisci le tue risorse

Questo tutorial include un AWS CloudFormation modello per una configurazione rapida. Puoi rivederlo e personalizzarlo in base alle tue esigenze. Il AWS CloudFormation modello genera le seguenti risorse:

- Utenti e politiche per:
  - DataLakeAdmin
  - DataAnalystUSA
  - DataAnalystJP
- Impostazioni e autorizzazioni del data lake Lake Formation
- Una funzione Lambda (per risorse AWS CloudFormation personalizzate supportate da Lambda) utilizzata per copiare file di dati di esempio dal bucket Amazon S3 pubblico al bucket Amazon S3
- Un bucket Amazon S3 che funge da data lake
- Un AWS Glue Data Catalog database, una tabella e una partizione

Crea le tue risorse

Segui questi passaggi per creare le tue risorse utilizzando il AWS CloudFormation modello.

1. Accedi alla AWS CloudFormation console all'indirizzo https://console.aws.amazon.com/cloudformation nella regione Stati Uniti orientali (Virginia settentrionale).
2. Scegli Launch Stack.
3. Scegli Avanti nella schermata Crea stack.
4. Inserisci un nome per lo stack.
5. Per DatalakeAdminUserNamee DatalakeAdminUserPassword, inserisci il nome utente e la password IAM per l'utente amministratore di data lake.
6. Per DataAnalystUsUserNamee DataAnalystUsUserPassword, inserisci il nome utente e la password per il nome utente e la password che desideri per l'utente analista di dati responsabile del mercato statunitense.

7. Per DataAnalystJpUserNamee DataAnalystJpUserPassword, inserisci il nome utente e la password per il nome utente e la password che desideri assegnare all'analista di dati responsabile del mercato giapponese.
8. Per DataLakeBucketName, inserisci il nome del tuo bucket di dati.
9. Per DatabaseName, e TableNamelascia come impostazione predefinita.
10. Seleziona Next (Successivo).
11. Nella pagina successiva, scegli Avanti.
12. Controlla i dettagli nella pagina finale e seleziona Riconosco che AWS CloudFormation potrebbe creare risorse IAM.
13. Seleziona Create (Crea).

    Il completamento della creazione dello stack può richiedere un minuto.

Fase 2: Interrogazione senza filtri di dati

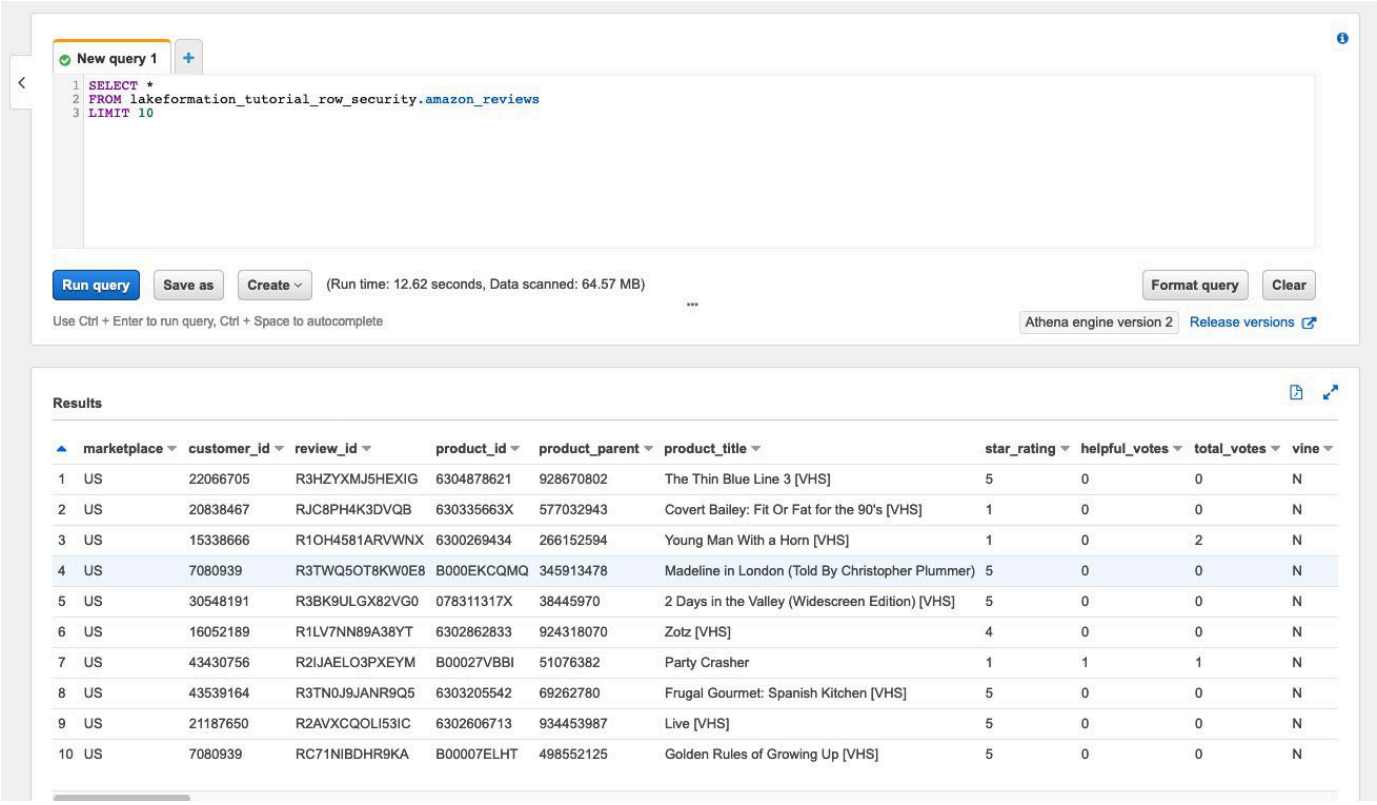
Dopo aver configurato l'ambiente, puoi interrogare la tabella delle recensioni dei prodotti. Per prima cosa interrogate la tabella senza controlli di accesso a livello di riga per assicurarvi di poter vedere i dati. Se esegui query in Amazon Athena per la prima volta, devi configurare la posizione dei risultati della query.

Interroga la tabella senza controllo degli accessi a livello di riga

1. Accedi alla Athena console all'indirizzo https://console.aws.amazon.com/athena/ come `DatalakeAdmin` utente ed esegui la seguente query:

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

   La schermata seguente mostra il risultato della query. Questa tabella ha una sola partizione`product_category=Video`, quindi ogni record è un commento di recensione per un prodotto video.

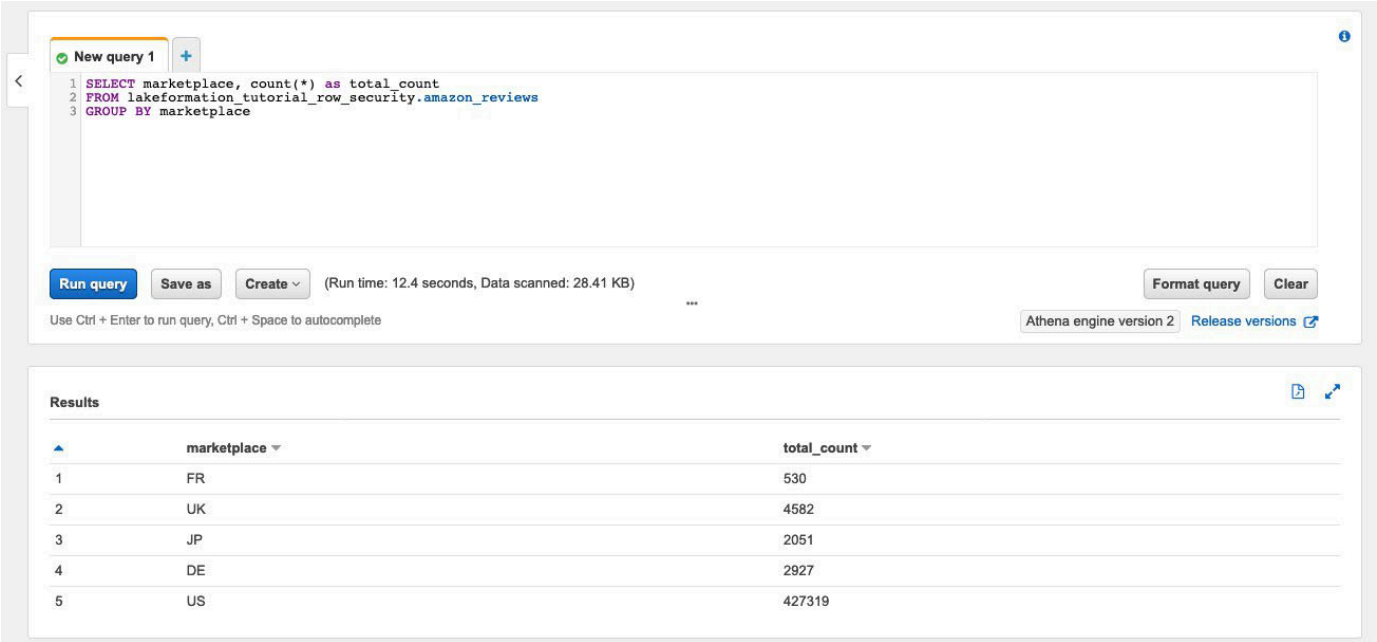


	marketplace	customer_id	review_id	product_id	product_parent	product_title	star_rating	helpful_votes	total_votes	vine
1	US	22066705	R3HZYXMJ5HEXIG	6304878621	928670802	The Thin Blue Line 3 [VHS]	5	0	0	N
2	US	20838467	RJC8PH4K3DVQB	630335663X	577032943	Covert Bailey: Fit Or Fat for the 90's [VHS]	1	0	0	N
3	US	15338666	R1OH4581ARVWNX	6300269434	266152594	Young Man With a Horn [VHS]	1	0	2	N
4	US	7080939	R3TWQ5OT8KW0E8	B000EKCQMQ	345913478	Madeline in London (Told By Christopher Plummer)	5	0	0	N
5	US	30548191	R3BK9ULGX82VG0	078311317X	38445970	2 Days in the Valley (Widescreen Edition) [VHS]	5	0	0	N
6	US	16052189	R1LV7NN89A38YT	6302862833	924318070	Zotz [VHS]	4	0	0	N
7	US	43430756	R2IJAELO3PXEYM	B00027VBBI	51076382	Party Crasher	1	1	1	N
8	US	43539164	R3TN0J9JANR9Q5	6303205542	69262780	Frugal Gourmet: Spanish Kitchen [VHS]	5	0	0	N
9	US	21187650	R2AVXCQOLI53IC	6302606713	934453987	Live [VHS]	5	0	0	N
10	US	7080939	RC71NIBDHR9KA	B00007ELHT	498552125	Golden Rules of Growing Up [VHS]	5	0	0	N

2. Quindi, esegui una query di aggregazione per recuperare il numero totale di record per. `marketplace`

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

La schermata seguente mostra il risultato della query. La `marketplace` colonna ha cinque valori diversi. Nei passaggi successivi, imposterai i filtri basati su righe utilizzando la `marketplace` colonna.



Passaggio 3: configura i filtri di dati e concedi le autorizzazioni

Questo tutorial utilizza due analisti di dati: uno responsabile del mercato statunitense e l'altro del mercato giapponese. Ogni analista utilizza Athena per analizzare le recensioni dei clienti solo per il proprio marketplace specifico. Crea due diversi filtri di dati, uno per l'analista responsabile del mercato statunitense e un altro per quello responsabile del mercato giapponese. Quindi, concedi agli analisti le rispettive autorizzazioni.

Crea filtri per i dati e concedi le autorizzazioni

1. Crea un filtro per limitare l'accesso ai US `marketplace` dati.

   a. Accedi alla console di Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/ nella regione Stati Uniti orientali (Virginia settentrionale) come `DatalakeAdmin` utente.

   b. Scegli Filtri dati.

   c. Scegli Crea nuovo filtro.

   d. Per il nome del filtro dati, inserisci`amazon_reviews_US`.

   e. Per Database Target, scegli il database`lakeformation_tutorial_row_security`.

   f. Per la tabella Target, scegli la tabella`amazon_reviews`.

   g. Per l'accesso a livello di colonna, lascia come impostazione predefinita.

h. Per Espressione del filtro Row, immettere. `marketplace='US'`

i. Scegli Create Filter (Crea filtro).

2. Create un filtro per limitare l'accesso ai `marketplace` dati giapponesi.

a. Nella pagina Filtri dati, scegli Crea nuovo filtro.

b. Per il nome del filtro dati, inserisci`amazon_reviews_JP`.

c. Per Database Target, scegli il database`lakeformation_tutorial_row_security`.

d. Per la tabella Target, scegli`table amazon_reviews`.

e. Per l'accesso a livello di colonna, lascia come impostazione predefinita.

f. Per Espressione del filtro Row, immettere. `marketplace='JP'`

g. Scegli Create Filter (Crea filtro).

3. Successivamente, concedi le autorizzazioni agli analisti di dati utilizzando questi filtri di dati. Segui questi passaggi per concedere le autorizzazioni all'analista di dati statunitense (): `DataAnalystUS`

a. In Autorizzazioni, scegli Autorizzazioni Data lake.

b. In Autorizzazione dati, scegli Concedi.

c. Per Principal, scegli utenti e ruoli IAM e seleziona il ruolo`DataAnalystUS`.

d. Per i tag LF o le risorse del catalogo, scegli Named data catalog resources.

e. Per Database, scegliere `lakeformation_tutorial_row_security`.

f. Per Tabelle (facoltativo), scegliete. `amazon_reviews`

g. Per i filtri di dati, facoltativo, seleziona. `amazon_reviews_US`

h. Per le autorizzazioni del filtro dati, seleziona Seleziona.

i. Scegli Concessione.

4. Segui questi passaggi per concedere le autorizzazioni all'analista di dati giapponese (): `DataAnalystJP`

a. In Autorizzazioni, scegli Autorizzazioni Data Lake.

b. In Autorizzazione dati, scegli Concedi.

c. Per Principal, scegli utenti e ruoli IAM e seleziona il ruolo`DataAnalystJP`.

d. Per i tag LF o le risorse del catalogo, scegli Named data catalog resources.

e. Per Database, scegliere `lakeformation_tutorial_row_security`.

f. Per Tabelle (facoltativo), scegliete. `amazon_reviews`

g. Per i filtri di dati, facoltativo, seleziona. `amazon_reviews_JP`

h. Per le autorizzazioni del filtro dati, seleziona Seleziona.

i. Scegli Concessione.

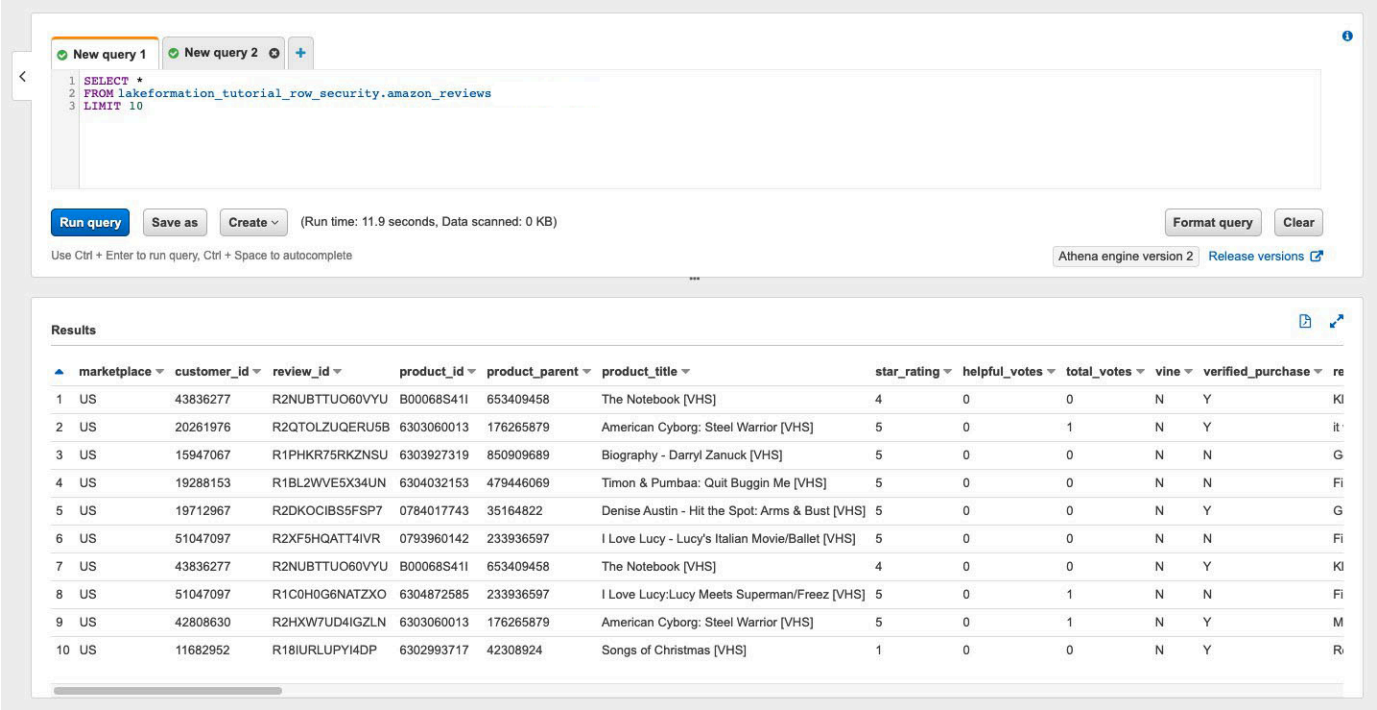
Passaggio 4: Interrogazione con filtri di dati

Con i filtri dei dati allegati alla tabella delle recensioni dei prodotti, esegui alcune query e scopri come vengono applicate le autorizzazioni da Lake Formation.

1. Accedi alla console Athena all'indirizzo https://console.aws.amazon.com/athena/ come utente. `DataAnalystUS`
2. Esegui la seguente query per recuperare alcuni record, che vengono filtrati in base alle autorizzazioni a livello di riga che abbiamo definito:

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

La schermata seguente mostra il risultato della query.



	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. Allo stesso modo, esegui una query per contare il numero totale di record per marketplace.

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

Il risultato della query mostra solo `marketplace` US i risultati. Questo perché all'utente è consentito visualizzare solo le righe in cui il valore della `marketplace` colonna è uguale aUS.

4. Passa all'`DataAnalystJP`utente ed esegui la stessa query.

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

Il risultato della query mostra che solo i record appartengono a `JPmarketplace`.

5. Esegui la query per contare il numero totale di record per`marketplace`.

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

Il risultato della query mostra solo la riga appartenente a `JPmarketplace`.

Fase 5: Pulire le AWS risorse

Pulizia delle risorse

Per evitare addebiti indesideratiAccount AWS, puoi eliminare le AWS risorse che hai utilizzato per questo tutorial.

- Elimina lo stack di formazione del cloud.

Condivisione di un data lake utilizzando il controllo degli accessi basato su tag Lake Formation e risorse denominate

Questo tutorial dimostra come configurare AWS Lake Formation la condivisione sicura dei dati archiviati all'interno di un data lake con più aziende, organizzazioni o unità aziendali, senza dover

copiare l'intero database. Esistono due opzioni per condividere database e tabelle con altri utenti Account AWS utilizzando il controllo degli accessi tra account di Lake Formation:

- Controllo degli accessi basato su tag Lake Formation (consigliato)

  Il controllo degli accessi basato su tag Lake Formation è una strategia di autorizzazione che definisce le autorizzazioni in base agli attributi. In Lake Formation, questi attributi sono chiamati LF-tag. Per ulteriori informazioni, consulta Gestione di un data lake utilizzando il controllo degli accessi basato su tag Lake Formation.

- Risorse denominate Lake Formation

  Il metodo delle risorse denominato Lake Formation è una strategia di autorizzazione che definisce le autorizzazioni per le risorse. Le risorse includono database, tabelle e colonne. Gli amministratori del Data Lake possono assegnare e revocare le autorizzazioni sulle risorse di Lake Formation. Per ulteriori informazioni, consulta Condivisione dei dati tra account in Lake Formation.

  Consigliamo di utilizzare risorse denominate se l'amministratore del data lake preferisce concedere le autorizzazioni in modo esplicito alle singole risorse. Quando utilizzi il metodo della risorsa denominata per concedere le autorizzazioni Lake Formation su una risorsa Data Catalog a un account esterno, Lake Formation utilizza AWS Resource Access Manager (AWS RAM) per condividere la risorsa.

Argomenti

Destinatari principali

Questo tutorial è destinato a amministratori di dati, ingegneri di dati e analisti di dati. Quando si tratta di condividere le tabelle di Data Catalog AWS Glue e amministrare le autorizzazioni in Lake Formation, gli amministratori dei dati all'interno degli account di produzione hanno la proprietà funzionale in base alle funzioni supportate e possono concedere l'accesso a vari consumatori, organizzazioni esterne e account. La tabella seguente elenca i ruoli utilizzati in questo tutorial:

Ruolo	Descrizione
DataLakeAdminProducer	L'utente IAM amministratore del data lake dispone dei seguenti accessi: • Accesso completo in lettura, scrittura e aggiornamento a tutte le risorse del Data Catalog • Possibilità di concedere autorizzazioni alle risorse • Può creare collegamenti a risorse per la tabella condivisa • Può allegare tag LF alle risorse, il che fornisce l'accesso ai principali in base a qualsiasi politica creata dai data steward
DataLakeAdminConsumer	L'utente IAM di data lake admin ha il seguente accesso: • Accesso completo in lettura, scrittura e aggiornamento a tutte le risorse del Data Catalog • Possibilità di concedere autorizzazioni alle risorse • Può creare collegamenti a risorse per la tabella condivisa • Può allegare tag LF alle risorse, il che fornisce l'accesso ai principali in base a qualsiasi politica creata dai data steward
DataAnalyst	L' DataAnalyst utente ha il seguente accesso:

Ruolo	Descrizione
	• Accesso granulare alle risorse condivise dalle politiche di accesso basate su tag di Lake Formation o utilizzando il metodo delle risorse denominate

Configura le impostazioni di Lake Formation Data Catalog nell'account produttore

Prima di iniziare questo tutorial, è necessario disporre di un Account AWS file da utilizzare per accedere come utente amministrativo con le autorizzazioni corrette. Per ulteriori informazioni, consulta Completa i processi di configurazione iniziali AWS.

Il tutorial presuppone che tu abbia familiarità con IAM. Per informazioni su IAM, consulta la IAM User Guide.

Configura le impostazioni di Lake Formation Data Catalog nell'account produttore

> **Note**
>
> In questo tutorial, l'account che ha la tabella di origine è chiamato account produttore e l'account che deve accedere alla tabella di origine è chiamato account consumatore.

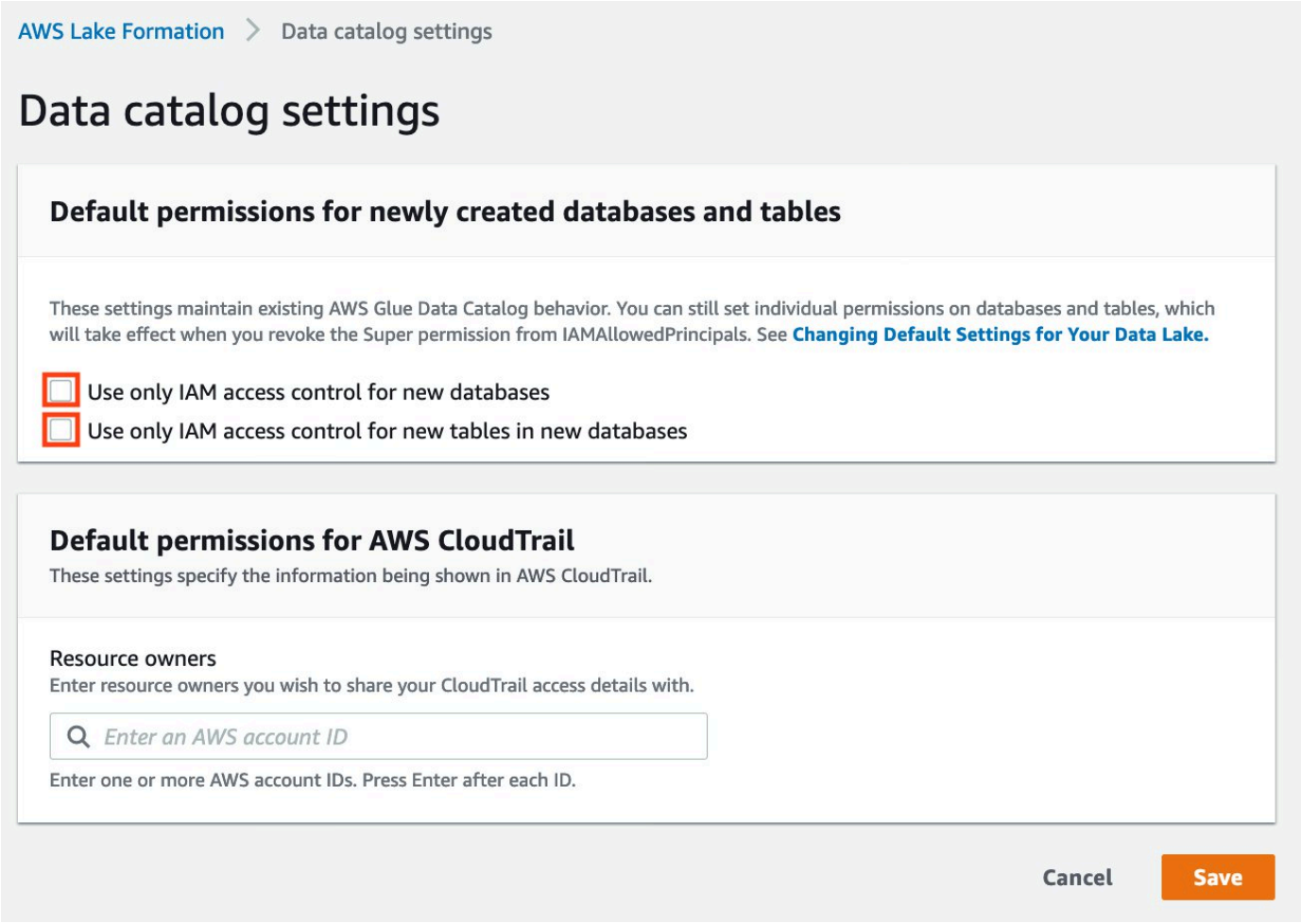
Lake Formation fornisce il proprio modello di gestione delle autorizzazioni. Per mantenere la retrocompatibilità con il modello di autorizzazione IAM, per impostazione predefinita l'`Super`autorizzazione viene concessa al gruppo `IAMAllowedPrincipals` su tutte AWS Glue Data Catalog le risorse esistenti. Inoltre, le impostazioni di controllo degli accessi Use only IAM sono abilitate per le nuove risorse del Data Catalog. Questo tutorial utilizza un controllo degli accessi granulare utilizzando le autorizzazioni di Lake Formation e utilizza le politiche IAM per un controllo degli accessi a grana grossolana. Per informazioni dettagliate, consulta Metodi per il controllo granulare degli accessi. Pertanto, prima di utilizzare un AWS CloudFormation modello per una configurazione rapida, è necessario modificare le impostazioni del Lake Formation Data Catalog nell'account produttore.

> **Important**
>
> Questa impostazione influisce su tutti i database e le tabelle appena creati, quindi consigliamo vivamente di completare questo tutorial in un account non di produzione o in un nuovo account. Inoltre, se utilizzi un account condiviso (ad esempio l'account di sviluppo della tua azienda), assicurati che ciò non influisca sulle risorse altrui. Se preferisci mantenere le impostazioni di sicurezza predefinite, devi completare un passaggio aggiuntivo durante la condivisione delle risorse con altri account, in cui revochi l'autorizzazione Super predefinita dal `IAMAllowedPrincipals` database o dalla tabella. Discuteremo i dettagli più avanti in questo tutorial.

Per configurare le impostazioni di Lake Formation Data Catalog nell'account produttore, completa i seguenti passaggi:

1. Accedi all'account AWS Management Console utilizzando l'account produttore come utente amministratore o come utente con l'autorizzazione dell'`PutDataLakeSettings`API Lake Formation.
2. Sulla console Lake Formation, nel pannello di navigazione, in Data Catalog, scegli Impostazioni.
3. Deseleziona Usa solo il controllo di accesso IAM per i nuovi database e Usa solo il controllo di accesso IAM per le nuove tabelle nei nuovi database

   Seleziona Salva.



Inoltre, puoi rimuovere `CREATE_DATABASE` le autorizzazioni per la `IAMAllowedPrincipals` sezione Ruoli e attività amministrative, Creatori di database. Solo allora, puoi decidere chi può creare un nuovo database tramite le autorizzazioni di Lake Formation.

Passaggio 1: Fornisci le tue risorse utilizzando modelli AWS CloudFormation

Il CloudFormation modello per l'account produttore genera le seguenti risorse:

- Un bucket Amazon S3 che funge da data lake.
- Una funzione Lambda (per risorse personalizzate supportate da Lambda). AWS CloudFormation Utilizziamo la funzione per copiare file di dati di esempio dal bucket Amazon S3 pubblico al tuo bucket Amazon S3.
- Utenti e politiche IAM:. DataLakeAdminProducer

- Le impostazioni e le autorizzazioni appropriate di Lake Formation, tra cui:
  - Definizione dell'amministratore del data lake di Lake Formation nell'account produttore
  - Registrazione di un bucket Amazon S3 come ubicazione del data lake Lake Formation (account produttore)
- Un AWS Glue Data Catalog database, una tabella e una partizione. Poiché esistono due opzioni per la condivisione delle risorseAccount AWS, questo modello crea due set separati di database e tabelle.

Il AWS CloudFormation modello per l'account consumatore genera le seguenti risorse:

- Utenti e politiche IAM:
  - DataLakeAdminConsumer
  - DataAnalyst
- Un database AWS Glue Data Catalog Questo database serve per creare collegamenti di risorse a risorse condivise.

Crea le tue risorse nell'account del produttore

1. Accedi alla AWS CloudFormation console all'indirizzo https://console.aws.amazon.com/cloudformation nella regione Stati Uniti orientali (Virginia settentrionale).
2. Scegli Launch Stack.
3. Seleziona Successivo.
4. Per il nome dello stack, inserisci un nome per lo stack, ad esempio. `stack-producer`
5. Nella sezione Configurazione utente, inserisci il nome utente e la password per `ProducerDatalakeAdminUserName` e. `ProducerDatalakeAdminUserPassword`
6. Per DataLakeBucketName, inserisci il nome del tuo bucket data lake. Questo nome deve essere univoco a livello globale.
7. Per DatabaseNamee TableName, lascia i valori predefiniti.
8. Seleziona Successivo.
9. Nella pagina successiva, scegli Avanti.
10. Controlla i dettagli nella pagina finale e seleziona Riconosco che AWS CloudFormation potrebbe creare risorse IAM.
11. Seleziona Create (Crea).

La creazione dello stack può richiedere fino a un minuto.

Crea le tue risorse nell'account consumatore

1. Accedi alla AWS CloudFormation console all'indirizzo https://console.aws.amazon.com/cloudformation nella regione Stati Uniti orientali (Virginia settentrionale).
2. Scegli Launch Stack.
3. Seleziona Successivo.
4. Per il nome dello stack, inserisci un nome per lo stack, ad esempio. `stack-consumer`
5. Nella sezione Configurazione utente, inserisci il nome utente e la password per `ConsumerDatalakeAdminUserName` e. `ConsumerDatalakeAdminUserPassword`
6. Per `DataAnalystUserName` e`DataAnalystUserPassword`, inserisci il nome utente e la password che desideri per l'utente IAM dell'analista di dati.
7. Per DataLakeBucketName, inserisci il nome del tuo bucket data lake. Questo nome deve essere univoco a livello globale.
8. Per DatabaseName, lascia i valori predefiniti.
9. Per`AthenaQueryResultS3BucketName`, inserisci il nome del bucket Amazon S3 in cui sono archiviati i risultati delle query di Amazon Athena. Se non ne hai uno, crea un bucket Amazon S3.
10. Seleziona Successivo.
11. Nella pagina successiva, scegli Avanti.
12. Controlla i dettagli nella pagina finale e seleziona Riconosco che AWS CloudFormation potrebbe creare risorse IAM.
13. Seleziona Create (Crea).

    La creazione dello stack può richiedere fino a un minuto.

> **Note**
>
> Dopo aver completato il tutorial, elimina lo stack AWS CloudFormation per evitare di incorrere in addebiti. Verifica che le risorse siano state eliminate correttamente nello stato dell'evento per lo stack.

Fase 2: Prerequisiti per la condivisione tra account di Lake Formation

Prima di condividere le risorse con Lake Formation, esistono dei prerequisiti sia per il metodo di controllo degli accessi basato su tag che per il metodo delle risorse denominate.

Prerequisiti completi per il controllo degli accessi basato su tag per la condivisione dei dati tra account

- Per ulteriori informazioni sui requisiti di condivisione dei dati tra account, consulta la Prerequisiti sezione del capitolo Condivisione dei dati tra account.

  Per condividere le risorse di Data Catalog con la versione 3 o successiva delle impostazioni della versione dell'account Cross, il concedente deve disporre delle autorizzazioni IAM definite nella policy AWS gestita del tuo account. `AWSLakeFormationCrossAccountManager`

  Se utilizzi la versione 1 o la versione 2 delle impostazioni della versione dell'account Cross, prima di poter utilizzare il metodo di controllo dell'accesso basato su tag per concedere l'accesso alle risorse tra account diversi, devi aggiungere il seguente oggetto di `JSON` autorizzazione alla politica delle risorse di Data Catalog nell'account produttore. Ciò concede all'account consumatore l'autorizzazione ad accedere al Data Catalog quando `glue:EvaluatedByLakeFormationTags` è vero. Inoltre, questa condizione diventa vera per le risorse a cui hai concesso l'autorizzazione utilizzando i tag di autorizzazione di Lake Formation sull'account del consumatore. Questa politica è necessaria per tutte le persone Account AWS a cui concedi le autorizzazioni.

  La seguente politica deve essere inclusa in un `Statement` elemento. Discuteremo la politica IAM completa nella prossima sezione.

```
{
    "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
        }
    }
}
```

Completa i prerequisiti per la condivisione tra account del metodo delle risorse denominate

1. Se nel tuo account non è presente alcuna politica sulle risorse di Data Catalog, il cross-account di Lake Formation ti consente di procedere come al solito. Tuttavia, se esiste una politica in materia di risorse di Data Catalog, devi aggiungere la seguente dichiarazione per consentire che le concessioni tra account abbiano esito positivo se effettuate con il metodo della risorsa denominata. Se prevedi di utilizzare solo il metodo della risorsa denominata o solo il metodo di controllo dell'accesso basato su tag, puoi saltare questo passaggio. In questo tutorial, valutiamo entrambi i metodi e dobbiamo aggiungere la seguente politica.

   La seguente politica deve trovarsi all'interno di un `Statement` elemento. Discuteremo la politica IAM completa nella prossima sezione.

   ```
   {
            "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. Quindi, aggiungi la politica AWS Glue Data Catalog delle risorse utilizzando AWS Command Line Interface (AWS CLI).

   Se concedi autorizzazioni per più account utilizzando sia il metodo di controllo degli accessi basato su tag che il metodo Named Resource, devi impostare l'`EnableHybrid`argomento su «true» quando aggiungi le politiche precedenti. Perché questa opzione non è attualmente supportata sulla console ed è necessario utilizzare l'API e. `glue:PutResourcePolicy` AWS CLI

   Innanzitutto, crea un documento di policy (come policy.json) e aggiungi le due politiche precedenti. *Sostituisci consumer-account-idcon l'ID dell'account di chi Account AWS riceve la sovvenzione, la regione con la regione del catalogo dati contenente i database e le tabelle per i quali stai concedendo le autorizzazioni e l'account-id con l'ID del produttore.* Account AWS

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": "ram.amazonaws.com"
            },
            "Action": "glue:ShareResource",
            "Resource": [
                "arn:aws:glue:region:account-id:table/*/*",
                "arn:aws:glue:region:account-id:database/*",
                "arn:aws:glue:region:account-id:catalog"
            ]
        },
        {
            "Effect": "Allow",
            "Principal": {
                "AWS": "region:account-id"
            },
            "Action": "glue:*",
            "Resource": [
                "arn:aws:glue:region:account-id:table/*/*",
                "arn:aws:glue:region:account-id:database/*",
                "arn:aws:glue:region:account-id:catalog"
```

```
            ],
            "Condition": {
                "Bool": {
                    "glue:EvaluatedByLakeFormationTags": "true"
                }
            }
        }
    ]
}
```

Immettete il seguente comando. AWS CLI Sostituisci *glue-resource-policy* con i valori corretti (ad esempio file: //policy.json).

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

Per ulteriori informazioni, vedere. put-resource-policy

Fase 3: Implementare la condivisione tra account utilizzando il metodo di controllo degli accessi basato su tag

In questa sezione, ti illustreremo i seguenti passaggi di alto livello:

1. Definisci un tag LF.
2. Assegna il tag LF alla risorsa di destinazione.
3. Concedi le autorizzazioni LF-Tag all'account consumatore.
4. Concedi le autorizzazioni relative ai dati all'account consumatore.
5. Facoltativamente, revoca le autorizzazioni per il database, le `IAMAllowedPrincipals` tabelle e le colonne.
6. Crea un link di risorsa alla tabella condivisa.
7. Crea un tag LF e assegnalo al database di destinazione.
8. Concedi le autorizzazioni relative ai dati LF-Tag all'account consumatore.

Definisci un tag LF

> **Note**
>
> Se hai effettuato l'accesso al tuo account produttore, esci prima di completare i seguenti passaggi.

1. Accedi all'account produttore come amministratore del data lake all'indirizzo https://console.aws.amazon.com/lakeformation/. Usa il numero di account del produttore, il nome utente IAM (l'impostazione predefinita è`DatalakeAdminProducer`) e la password che hai specificato durante la creazione AWS CloudFormation dello stack.
2. Nella console di Lake Formation (https://console.aws.amazon.com/lakeformation/), nel pannello di navigazione, in Autorizzazioni e in Ruoli e attività amministrativi, scegli LF-Tags.
3. Scegliete Aggiungi LF-tag.

Assegna il tag LF alla risorsa di destinazione

Assegna il tag LF alla risorsa di destinazione e concedi le autorizzazioni relative ai dati a un altro account

In qualità di amministratore del data lake, puoi allegare tag alle risorse. Se prevedi di utilizzare un ruolo separato, potresti dover concedere autorizzazioni di descrizione e allegare al ruolo separato.

1. Nel riquadro di navigazione, in Data Catalog, seleziona Database.
2. Seleziona il database di destinazione (`lakeformation_tutorial_cross_account_database_tbac`) e nel menu Azioni, scegli Modifica tag LF.

   In questo tutorial, puoi assegnare un tag LF a un database, ma puoi anche assegnare tag LF a tabelle e colonne.
3. Scegliete Assegna nuovo LF-tag.
4. Aggiungi la chiave e il valore`Confidentiality`. `public`
5. Seleziona Salva.

Concedi l'autorizzazione LF-Tag all'account consumatore

Sempre nell'account produttore, concedi le autorizzazioni all'account consumer per accedere all'LF-Tag.

1. Nel pannello di navigazione, in Autorizzazioni, Ruoli e attività amministrativi, Autorizzazioni LF-Tag, scegli Concedi.
2. Per Principali, scegli Account esterni.
3. Inserisci l'Account AWSID di destinazione.

   Account AWSall'interno della stessa organizzazione appaiono automaticamente. Altrimenti, devi inserire manualmente l'Account AWSID. Al momento della stesura di questo documento, il controllo degli accessi basato su tag Lake Formation non supporta la concessione di autorizzazioni a organizzazioni o unità organizzative.
4. Per i tag LF, scegliete la chiave e i valori del tag LF da condividere con l'account consumatore (chiave e valore). **`Confidentiality`** `public`
5. Per Autorizzazioni, seleziona Descrivi per le autorizzazioni LF-Tag.

   Le autorizzazioni LF-Tag sono autorizzazioni concesse all'account consumatore. Le autorizzazioni concedibili sono autorizzazioni che l'account consumatore può concedere ad altri destinatari.
6. Scegli Concessione.

   A questo punto, l'amministratore del consumer data lake dovrebbe essere in grado di trovare il policy tag condiviso tramite la console dell'account consumer Lake Formation, in Autorizzazioni, ruoli e attività amministrativi, LF-Tags.

Concedi l'autorizzazione ai dati all'account consumatore

Ora forniremo l'accesso ai dati all'account consumatore specificando un'espressione LF-Tag e concedendo all'account consumatore l'accesso a qualsiasi tabella o database che corrisponda all'espressione.

1. Nel pannello di navigazione, in Autorizzazioni, Autorizzazioni Data Lake, scegli Concedi.
2. Per Principali, scegli Account esterni e inserisci l'ID di destinazione. Account AWS
3. Per i tag LF o le risorse del catalogo, scegli la chiave e i valori del tag LF da condividere con l'account consumatore (chiave e valore). **`Confidentiality`** `public`

4. Per Autorizzazioni, in Risorse abbinate ai tag LF (consigliato) scegli Aggiungi tag LF.
5. Seleziona la chiave e il valore del tag che viene condiviso con l'account consumatore (chiave e valore). `Confidentiality public`
6. Per le autorizzazioni del database, seleziona Descrivi in Autorizzazioni del database per concedere le autorizzazioni di accesso a livello di database.
7. L'amministratore del consumer data lake dovrebbe essere in grado di trovare il policy tag condiviso tramite l'account consumer sulla console di Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/, in Autorizzazioni, ruoli e attività amministrativi, LF-Tags.
8. Seleziona Descrivi in Autorizzazioni concesse in modo che l'account consumer possa concedere autorizzazioni a livello di database ai suoi utenti.
9. Per le autorizzazioni relative alle tabelle e alle colonne, seleziona Seleziona e descrivi in Autorizzazioni per le tabelle.
10. Seleziona Seleziona e descrivi in Autorizzazioni concedibili.
11. Scegli Concessione.

Revoca l'autorizzazione per **`IAMAllowedPrincipals`** il database, le tabelle e le colonne (facoltativo).

All'inizio di questo tutorial, hai modificato le impostazioni del Lake Formation Data Catalog. Se hai saltato quella parte, questo passaggio è obbligatorio. Se hai modificato le impostazioni del Lake Formation Data Catalog, puoi saltare questo passaggio.

In questo passaggio, dobbiamo revocare l'autorizzazione Super predefinita dal database o `IAMAllowedPrincipals` dalla tabella. Per informazioni dettagliate, consulta Passaggio 4: Passa i tuoi archivi dati al modello di autorizzazioni Lake Formation.

Prima di revocare l'autorizzazione`IAMAllowedPrincipals`, assicurati di aver concesso ai principali IAM esistenti l'autorizzazione necessaria tramite Lake Formation. Ciò include tre passaggi:

1. Aggiungi l'autorizzazione IAM all'utente o al ruolo IAM di destinazione con l'`GetDataAccess`azione Lake Formation (con policy IAM).
2. Concedi all'utente o al ruolo IAM di destinazione le autorizzazioni relative ai dati di Lake Formation (modifica, selezione e così via).

3. Quindi, revoca le autorizzazioni per. `IAMAllowedPrincipals` Altrimenti, dopo aver revocato le autorizzazioni per`IAMAllowedPrincipals`, i principali IAM esistenti potrebbero non essere più in grado di accedere al database di destinazione o al Data Catalog.

   La revoca dell'autorizzazione Super for `IAMAllowedPrincipals` è necessaria quando si desidera applicare il modello di autorizzazione Lake Formation (anziché il modello di policy IAM) per gestire l'accesso degli utenti all'interno di un singolo account o tra più account utilizzando il modello di autorizzazione Lake Formation. Non è necessario revocare l'autorizzazione `IAMAllowedPrincipals` per altre tabelle in cui si desidera mantenere il modello di policy IAM tradizionale.

   A questo punto, l'amministratore di Data Lake dell'account consumer dovrebbe essere in grado di trovare il database e la tabella condivisi tramite l'account consumer sulla console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/, in Data Catalog, database. In caso contrario, verifica se quanto segue è configurato correttamente:

   1. Il tag di policy e i valori corretti vengono assegnati ai database e alle tabelle di destinazione.
   2. L'autorizzazione corretta per il tag e l'autorizzazione per i dati vengono assegnate all'account consumatore.
   3. Revoca la super autorizzazione predefinita dal `IAMAllowedPrincipals` database o dalla tabella.

Crea un link di risorsa alla tabella condivisa

Quando una risorsa viene condivisa tra account e le risorse condivise non vengono inserite nel catalogo dati degli account consumatori. Per renderli disponibili e interrogare i dati sottostanti di una tabella condivisa utilizzando servizi come Athena, dobbiamo creare un collegamento di risorsa alla tabella condivisa. Un link a una risorsa è un oggetto del Data Catalog che è un collegamento a un database o a una tabella locale o condivisa. Per informazioni dettagliate, consultare Creazione di collegamenti alle risorse. Creando un collegamento a una risorsa, puoi:

- Assegna un nome diverso a un database o a una tabella in linea con le politiche di denominazione delle risorse del Data Catalog.
- Utilizza servizi come Athena e Redshift Spectrum per interrogare database o tabelle condivisi.

Per creare un collegamento a una risorsa, completa i seguenti passaggi:

1. Se hai effettuato l'accesso al tuo account consumatore, esci.
2. Accedi come amministratore di Data Lake dell'account consumer. Usa l'ID dell'account consumer, il nome utente IAM (predefinito DatalakeAdminConsumer) e la password che hai specificato durante la creazione AWS CloudFormation dello stack.
3. Sulla console Lake Formation (https://console.aws.amazon.com/lakeformation/), nel riquadro di navigazione, in Data Catalog, Databases, scegli il database condivisolakeformation_tutorial_cross_account_database_tbac.

   Se non vedi il database, rivedi i passaggi precedenti per vedere se tutto è configurato correttamente.
4. Scegli Visualizza tabelle.
5. Scegli la tabella condivisaamazon_reviews_table_tbac.
6. Nel menu Azioni, scegli Crea link a una risorsa.
7. Per il nome del link alla risorsa, inserisci un nome (per questo tutorial,amazon_reviews_table_tbac_resource_link).
8. In Database, seleziona il database in cui viene creato il collegamento alla risorsa (per questo post, lo stack AWS CloudFormation n ha creato il databaselakeformation_tutorial_cross_account_database_consumer).
9. Seleziona Create (Crea).

   Il link alla risorsa viene visualizzato in Catalogo dati, Tabelle.

Create un tag LF e assegnatelo al database di destinazione

I tag Lake Formation si trovano nello stesso Data Catalog delle risorse. Ciò significa che i tag creati nell'account produttore non possono essere utilizzati quando si concede l'accesso ai link alle risorse nell'account consumatore. È necessario creare un set separato di tag LF nell'account consumatore per utilizzare il controllo degli accessi basato sui tag LF quando si condividono i link alle risorse nell'account consumatore.

1. Definisci il tag LF nell'account consumatore. Per questo tutorial, utilizziamo chiavi `Division` e valori `sales` e`marketing`. `analyst`
2. Assegna la chiave `Division` e il valore del tag LF `analyst` al databaselakeformation_tutorial_cross_account_database_consumer, dove viene creato il collegamento alla risorsa.

Concedi l'autorizzazione ai dati LF-Tag al consumatore

Come ultimo passaggio, concedi l'autorizzazione ai dati di LF-Tag al consumatore.

1. Nel pannello di navigazione, in Autorizzazioni, Autorizzazioni Data lake, scegli Concedi.
2. Per Principal, scegli utenti e ruoli IAM e scegli l'utente. `DataAnalyst`
3. Per i tag LF o le risorse del catalogo, scegli Risorse abbinate ai tag LF (consigliato).
4. Scegliete Key Division and Value Analyst.
5. Per le autorizzazioni del database, seleziona Descrivi in Autorizzazioni del database.
6. Per le autorizzazioni per tabelle e colonne, seleziona Seleziona e descrivi in Autorizzazioni per le tabelle.
7. Scegli Concessione.
8. Ripeti questi passaggi per l'utente`DataAnalyst`, dove si trova la chiave LF-Tag e il valore è`Confidentiality`. `public`

   A questo punto, l'utente analista di dati nell'account consumatore dovrebbe essere in grado di trovare il database e il collegamento alla risorsa e di interrogare la tabella condivisa tramite la console Athena all'indirizzo https://console.aws.amazon.com/athena/. In caso contrario, verifica se quanto segue è configurato correttamente:

   - Il link alla risorsa viene creato per la tabella condivisa
   - Hai concesso all'utente l'accesso al tag LF condiviso dall'account produttore
   - Hai concesso all'utente l'accesso al tag LF associato al link alla risorsa e al database in cui è stato creato il link alla risorsa
   - Controlla di aver assegnato il tag LF corretto al link alla risorsa e al database in cui è stato creato il link alla risorsa

Fase 4: Implementazione del metodo di risorsa denominato

Per utilizzare il metodo della risorsa denominata, ti spieghiamo i seguenti passaggi di alto livello:

1. Facoltativamente, revoca l'autorizzazione per il database, `IAMAllowedPrincipals` le tabelle e le colonne.
2. Concedi l'autorizzazione ai dati all'account consumatore.
3. Accetta una condivisione di risorse daAWS Resource Access Manager.

4. Crea un link alla risorsa per la tabella condivisa.
5. Concedi al consumatore l'autorizzazione ai dati per la tabella condivisa.
6. Concedi al consumatore l'autorizzazione ai dati per il collegamento alla risorsa.

Revoca l'autorizzazione per **`IAMAllowedPrincipals`** il database, le tabelle e le colonne (facoltativo)

- All'inizio di questo tutorial, abbiamo modificato le impostazioni del Lake Formation Data Catalog. Se hai saltato quella parte, questo passaggio è obbligatorio. Per istruzioni, consulta il passaggio facoltativo nella sezione precedente.

Concedi l'autorizzazione ai dati all'account consumatore

1.
> **Note**
>
> Se hai effettuato l'accesso all'account produttore come altro utente, esci prima.

   Accedi alla console di Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/ utilizzando l'account producer data lake administrator utilizzando l'Account AWSID, il nome utente IAM (l'impostazione predefinita è`DatalakeAdminProducer`) e la password specificati durante la creazione AWS CloudFormation dello stack.
2. Nella pagina Autorizzazioni, in Autorizzazioni Data lake scegli Concedi.
3. In Principali, scegli Account esterni e inserisci uno o più Account AWS ID o AWS ID di organizzazioni. Per ulteriori informazioni, vedere: AWSOrganizations.

   Le organizzazioni a cui appartiene l'account produttore e che Account AWS fanno parte della stessa organizzazione vengono visualizzate automaticamente. Altrimenti, inserisci manualmente l'ID dell'account o l'ID dell'organizzazione.
4. Per i tag LF o le risorse del catalogo, scegliete. `Named data catalog resources`
5. In Database, scegliete il database. `lakeformation_tutorial_cross_account_database_named_resource`
6. Scegli Aggiungi tag LF.
7. In Tabelle, scegli Tutte le tabelle.

8. Per le autorizzazioni per le colonne della tabella, scegli Seleziona e Descrivi in Autorizzazioni per la tabella.
9. Seleziona Seleziona e descrivi, in Autorizzazioni concedibili.
10. Facoltativamente, per le autorizzazioni relative ai dati, scegli Accesso semplice basato su colonne se è richiesta la gestione delle autorizzazioni a livello di colonna.
11. Scegli Concessione.

Se non hai revocato l'autorizzazione per**`IAMAllowedPrincipals`**, viene visualizzato l'errore Grant permissions failed. A questo punto, dovresti vedere la tabella di destinazione condivisa AWS RAM con l'account consumatore in Autorizzazioni, Autorizzazioni dati.

Accetta una condivisione di risorse da AWS RAM

> **Note**
>
> Questo passaggio è necessario solo per la condivisione Account AWS basata, non per la condivisione basata sull'organizzazione.

1. Accedi alla AWS console all'indirizzo https://console.aws.amazon.com/connect/ utilizzando l'amministratore del data lake dell'account consumer utilizzando il nome utente IAM (l'impostazione predefinita è DatalakeAdminConsumer) e la password specificati durante la creazione AWS CloudFormation dello stack.
2. Sulla AWS RAM console, nel pannello di navigazione, sotto Shared with me, Resource shares, scegli la risorsa condivisa Lake Formation. Lo stato dovrebbe essere In sospeso.
3. Scegli Azione e Concedi.
4. Conferma i dettagli della risorsa e scegli Accetta condivisione delle risorse.

   A questo punto, l'amministratore del data lake dell'account consumer dovrebbe essere in grado di trovare la risorsa condivisa sulla console Lake Formation (https://console.aws.amazon.com/lakeformation/) in Data Catalog, Databases.

Crea un link alla risorsa per la tabella condivisa

- Segui le istruzioni riportate nel Fase 3: Implementare la condivisione tra account utilizzando il metodo di controllo degli accessi basato su tag passaggio 6 per

creare un link a una risorsa per una tabella condivisa. Assegna un nome al link alla risorsa`amazon_reviews_table_named_resource_resource_link`. Crea il link alla risorsa nel database`lakeformation_tutorial_cross_account_database_consumer`.

Concedi al consumatore l'autorizzazione ai dati per la tabella condivisa

Per concedere al consumatore l'autorizzazione ai dati per la tabella condivisa, completa i seguenti passaggi:

1. Nella console di Lake Formation (https://console.aws.amazon.com/lakeformation/), in Autorizzazioni, Autorizzazioni Data lake, scegli Concedi.
2. Per Principal, scegli utenti e ruoli IAM e scegli l'utente. `DataAnalyst`
3. Per i tag LF o le risorse del catalogo, scegli Risorse del catalogo dati denominato.
4. In Database, scegliete il database. `lakeformation_tutorial_cross_account_database_named_resource` Se non vedi il database nell'elenco a discesa, scegli Carica altro.
5. In Tabelle, scegli la tabella. `amazon_reviews_table_named_resource`
6. Per le autorizzazioni per tabelle e colonne, seleziona Seleziona e descrivi in Autorizzazioni per le tabelle.
7. Scegli Concessione.

Concedi al consumatore l'autorizzazione ai dati per il collegamento alla risorsa

Oltre a concedere all'utente del data lake l'autorizzazione ad accedere alla tabella condivisa, è necessario concedere all'utente del data lake l'autorizzazione ad accedere al collegamento alla risorsa.

1. Nella console Lake Formation (https://console.aws.amazon.com/lakeformation/), in Autorizzazioni, Autorizzazioni Data lake, scegli Concedi.
2. Per Principal, scegli utenti e ruoli IAM e scegli l'utente. `DataAnalyst`
3. Per i tag LF o le risorse del catalogo, scegli Risorse del catalogo dati denominato.
4. In Database, scegliete il database. `lakeformation_tutorial_cross_account_database_consumer` Se non vedi il database nell'elenco a discesa, scegli Carica altro.
5. In Tabelle, scegli la tabella. `amazon_reviews_table_named_resource_resource_link`

6. Per le autorizzazioni relative ai collegamenti alle risorse, seleziona Descrivi in Autorizzazioni per i collegamenti alle risorse.
7. Scegli Concessione.

   A questo punto, l'utente analista di dati nell'account consumatore dovrebbe essere in grado di trovare il database e il collegamento alla risorsa e di interrogare la tabella condivisa tramite la console Athena.

   In caso contrario, verifica se quanto segue è configurato correttamente:

   - Il link alla risorsa viene creato per la tabella condivisa
   - Hai concesso all'utente l'accesso alla tabella condivisa dall'account produttore
   - Hai concesso all'utente l'accesso al collegamento alla risorsa e al database per cui è stato creato il collegamento alla risorsa

Passaggio 5: Pulisci AWS le risorse

Per evitare addebiti indesideratiAccount AWS, puoi eliminare le AWS risorse che hai utilizzato per questo tutorial.

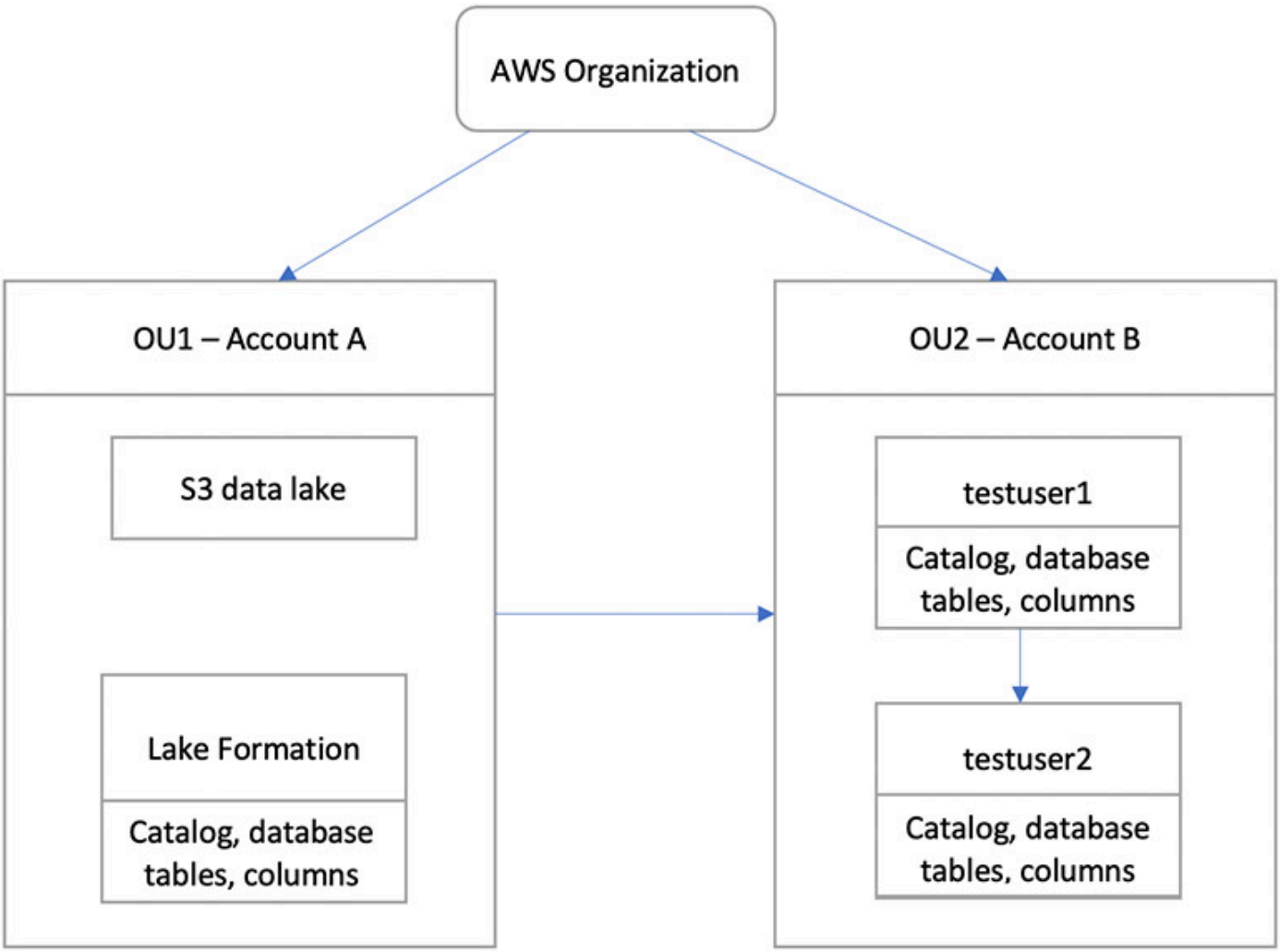
1. Accedi alla console di Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/ utilizzando l'account produttore ed elimina o modifica quanto segue:
   - AWS Resource Access Managercondivisione di risorse
   - Etichette di Lake Formation
   - AWS CloudFormation Stack
   - Impostazioni Lake Formation
   - AWS Glue Data Catalog
2. Accedi alla console di Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/ utilizzando l'account utente ed elimina o modifica quanto segue:
   - Etichette di Lake Formation
   - AWS CloudFormation Stack

Condivisione di un data lake utilizzando il controllo granulare degli accessi di un data lake

Questo tutorial fornisce step-by-step istruzioni su come condividere rapidamente e facilmente set di dati utilizzando Lake Formation quando si gestiscono più set di datiAccount AWS conAWS Organizations. Puoi definire autorizzazioni granulare per controllare l'accesso a dati sensibili.

Le procedure seguenti mostrano anche come un amministratore di data lake dell'account A può fornire un accesso granulare per l'account B e come un utente dell'account B, in qualità di amministratore dei dati, può concedere l'accesso granulare alla tabella condivisa ad altri utenti del proprio account. I data steward all'interno di ogni account possono delegare in modo indipendente l'accesso ai propri utenti, dando autonomia a ciascun team o linea di business (LOB).

Il caso d'uso presuppone che tu stia utilizzandoAWS Organizations per gestire il tuoAccount AWS. L'utente dell'Account A in un'unità organizzativa (OU1) concede l'accesso agli utenti dell'Account B in OU2. È possibile utilizzare lo stesso approccio quando non si utilizzano Organizations, ad esempio quando si hanno solo pochi account. Nel diagramma seguente è possibile accedere a un controllo granulare degli accessi dei set di un data lake. Il data lake è disponibile nell'Account A. L'amministratore del data lake dell'Account A fornisce un accesso dettagliato per l'Account B. Il diagramma mostra anche che un utente dell'Account B fornisce l'accesso a livello di colonna della tabella del data lake Account A a un altro utente nell'Account B.



Argomenti

Destinatari principali

Questo tutorial è destinato ai data steward, agli ingegneri dei dati e agli analisti di dati. Nella tabella seguente sono elencati i ruoli utilizzati in questo tutorial:

Ruolo	Descrizione
Amministratore IAM	Utente che ha la politicaAWS gestita:`AdministratorAccess` .
Amministrazione di un data lake	Utente che ha la policyAWS gestita:`AWSLakeFormationDataAdmin` associato al ruolo.
Analista dei dati	Utente che ha la politicaAWS gestita:`AmazonAthenaFullAccess` allegato.

Prerequisiti

Prima di iniziare questo tutorial, devi disporre di un alAccount AWS quale puoi eseguire l'accesso come utente amministrativo con le corrette autorizzazioni. Per ulteriori informazioni, consulta Completa i processi di configurazione iniziali AWS.

Il tutorial presuppone che tu abbia familiarità con IAM. Per informazioni su IAM, consulta la Guida per l'utente di IAM.

Per questo tutorial:

- Due unità organizzative:
  - OU1: contiene l'account A
  - OU2: contiene l'account B
- Una posizione (bucket) di un data lake Amazon S3 nell'account A.
- Un utente amministratore del data lake nell'Account A. È possibile creare un amministratore del data lake utilizzando la console Lake Formation (https://console.aws.amazon.com/lakeformation/) o ilPutDataLakeSettings funzionamento dell'API Lake Formation.
- Lake Formation configurato nell'account A e la posizione del data lake Amazon S3 registrata con Lake Formation nell'account A.
- Due utenti nell'Account B con le seguenti politiche gestite da IAM:
  - testuser1: ha le politicheAWS gestiteAWSLakeFormationDataAdmin allegate.

  - testuser2 — Ha la policyAWS gestita`AmazonAthenaFullAccess` allegata.
- Un database testdb nel database Lake Formation per l'account B.

Fase 1: fornire un account granulare degli accessi a un account

Scopri come un amministratore di data lake dell'account A fornisce un accesso granulare per l'account B.

Consegnazione dell'accesso granulare di un account

1. AccediAWS Management Console all'indirizzo https://console.aws.amazon.com/connect/ nell'Account A come amministratore del data lake.
2. Apri la console di Lake Formation (https://console.aws.amazon.com/lakeformation/) e scegli Inizia.
3. nel riquadro di navigazione, scegliere Database.
4. Scegliere Create database (Crea database).
5. Nella sezione Dettagli del database, seleziona Database.
6. Per Nome, inserisci un nome (per questo tutorial, usiamo`sampledb01`).
7. Assicurati che l'opzione Usa solo il controllo di accesso IAM per le nuove tabelle in questo database non sia selezionata. Lasciando questa opzione deselezionata possiamo controllare l'accesso da Lake Formation.
8. Scegliere Crea database.
9. Nella pagina Database, scegli il tuo database`sampledb01`.
10. Nel menu Azioni, scegli Concedi.
11. Nella sezione Concedi autorizzazioni, seleziona Account esterno.
12. ComeAccount AWS ID o IDAWS dell'organizzazione, inserisci l'ID dell'account B in OU2.
13. Per Tabella, scegli la tabella a cui desideri che l'Account B abbia accesso (per questo post, utilizziamo la tabella`acc_a_area`). Facoltativamente, puoi concedere l'accesso alle colonne all'interno della tabella, cosa che facciamo in questo post.
14. Per le colonne Includi, scegli le colonne a cui vuoi che l'Account B abbia accesso (per questo post, concediamo le autorizzazioni per il tipo, il nome e gli identificatori).
15. Per Colonne, scegli Includi colonne.
16. Per le autorizzazioni relative alla tabella, seleziona Seleziona.

17. Per le autorizzazioni concesse, seleziona Seleziona. Le autorizzazioni concesse sono necessarie affinché gli utenti amministratori dell'Account B possano concedere le autorizzazioni ad altri utenti nell'Account B.
18. Scegli Concessione.
19. Nel pannello di navigazione, seleziona Tabelle.
20. Potresti vedere una connessione attiva nella sezioneAccount AWS eAWS organizzazioni con accesso.

Creazione di un collegamento

I servizi integrati come Amazon Athena non possono accedere direttamente a database o tabelle tra account. Quindi, è necessario creare un collegamento alle risorse in modo che Athena possa accedere ai collegamenti alle risorse del tuo account ai database e alle tabelle di altri account. Crea un collegamento alla risorsa alla tabella (`acc_a_area`) in modo che gli utenti dell'Account B possano interrogare i suoi dati con Athena.

1. Accedi allaAWS console all'indirizzo https://console.aws.amazon.com/connect/ nell'account B come`testuser1`.
2. Nella console di Lake Formation (https://console.aws.amazon.com/lakeformation/), nel riquadro di navigazione, scegli Tabelle. Dovresti vedere le tabelle a cui l'Account A ha fornito l'accesso.
3. Seleziona la tabella `acc_a_area`.
4. Nel menu Azioni, scegli Crea link alla risorsa.
5. Per il nome del collegamento alla risorsa, inserisci un nome (per questo tutorial,`acc_a_area_rl`).
6. Per Database, scegli il tuo database (`testdb`).
7. Seleziona Create (Crea).
8. Nel pannello di navigazione, seleziona Tabelle.
9. Seleziona la tabella `acc_b_area_rl`.
10. Nel menu Azioni, scegli Visualizza dati.

    Verrai reindirizzato alla console Athena, dove dovresti vedere il database e la tabella.

    Ora puoi eseguire una query sulla tabella per vedere il valore della colonna per il quale è stato fornito l'accesso a testuser1 dall'Account B.

Passaggio 2: fornire un accesso dettagliato a un utente nello stesso account

Questa sezione mostra come un utente dell'Account B (`testuser1`), in qualità di amministratore dei dati, fornisce un accesso dettagliato a un altro utente dello stesso account (`testuser2`) al nome della colonna nella tabella condivisa`aac_b_area_rl`.

Consegnazione dell'accesso granulare degli utenti dello stesso account

1. Accedi allaAWS console all'indirizzo https://console.aws.amazon.com/connect/ nell'account B come`testuser1`.
2. Nella console di Lake Formation, nel riquadro di navigazione, scegliere Tabelle.

   Puoi concedere le autorizzazioni su una tabella tramite il collegamento alla risorsa. A tale scopo, nella pagina Tabelle, seleziona il link alla risorsa `eacc_b_area_rl`, nel menu Azioni, scegli Concedi sull'obiettivo.
3. Nella sezione Concedi le autorizzazioni, seleziona Il mio account.
4. Per gli utenti e i ruoli IAM, scegli l'utente`testuser2`.
5. In Colonna, scegli il nome della colonna.
6. Per le autorizzazioni relative alla tabella, seleziona Seleziona.
7. Scegli Concessione.

   Quando crei un link a una risorsa, solo tu puoi visualizzarlo e accedervi. Per consentire agli altri utenti del tuo account di accedere al link alla risorsa, devi concedere le autorizzazioni sul link alla risorsa stesso. È necessario concedere le autorizzazioni DESCRIVE o DROP. Nella pagina Tabelle, seleziona nuovamente la tua tabella e nel menu Azioni, scegli Concedi.
8. Nella sezione Concedi le autorizzazioni, seleziona Il mio account.
9. Per gli utenti e i ruoli IAM, seleziona l'utente`testuser2`.
10. Per le autorizzazioni relative al collegamento alla risorsa, seleziona Descrivi.
11. Scegli Concessione.
12. Accedi allaAWS console nell'Account B come`testuser2`.

    Sulla console Athena (https://console.aws.amazon.com/athena/), dovresti vedere il database e la tabella`aacc_b_area_rl`. Ora puoi eseguire una query sulla tabella per vedere il valore della colonna a cui`testuser2` ha accesso.

Accedere ai permessi di Lake Formation

AWS Lake Formationutilizza il AWS Glue Data Catalog per archiviare i metadati per i dati di Amazon S3 sotto forma di database e tabelle. Le tabelle memorizzano informazioni sui dati sottostanti, tra cui informazioni sullo schema, sulle partizioni e sulla posizione dei dati. I database sono raccolte di tabelle. Il Data Catalog contiene anche collegamenti a risorse, che sono collegamenti a database e tabelle condivisi in account esterni e vengono utilizzati per l'accesso tra account diversi ai dati nel data lake. Ogni AWS account dispone di un catalogo dati per AWS regione.

Lake Formation fornisce un modello di autorizzazioni del sistema di gestione dei database relazionali (RDBMS) per concedere o revocare l'accesso a database, tabelle e colonne nel Data Catalog con i dati sottostanti in Amazon S3.

Prima di conoscere i dettagli del modello di autorizzazioni di Lake Formation, è utile esaminare le seguenti informazioni di base:

- I data lake gestiti da Lake Formation risiedono in luoghi designati in Amazon Simple Storage Service (Amazon S3).
- Lake Formation gestisce un Data Catalog che contiene metadati sui dati di origine da importare nei tuoi data lake, come i dati nei log e nei database relazionali, e sui dati nei tuoi data lake in Amazon S3. I metadati sono organizzati come database e tabelle. Le tabelle di metadati contengono schema, posizione, partizionamento e altre informazioni sui dati che rappresentano. I database di metadati sono raccolte di tabelle.
- Il Lake Formation Data Catalog è lo stesso Data Catalog utilizzato daAWS Glue. Puoi utilizzare AWS Glue i crawler per creare tabelle del Data Catalog e puoi utilizzare i processi di AWS Glue estrazione, trasformazione e caricamento (ETL) per popolare i dati sottostanti nei tuoi data lake.
- I database e le tabelle del Data Catalog sono denominati risorse del Data Catalog. Le tabelle nel catalogo dati vengono chiamate tabelle di metadati per distinguerle dalle tabelle nelle fonti di dati o dai dati tabulari in Amazon S3. I dati a cui puntano le tabelle di metadati in Amazon S3 o nelle fonti di dati vengono definiti dati sottostanti.
- Un principale è un utente o un ruolo, un QuickSight utente o gruppo Amazon, un utente o un gruppo che si autentica con Lake Formation tramite un provider SAML o, per il controllo degli accessi tra account, ID AWS account, ID organizzazione o ID unità organizzativa.
- AWS Gluei crawler creano tabelle di metadati, ma puoi anche creare manualmente tabelle di metadati con la console Lake Formation, l'API o (). AWS Command Line Interface AWS CLI Quando crei una tabella di metadati, devi specificare una posizione. Quando si crea un database,

la posizione è facoltativa. Le posizioni delle tabelle possono essere posizioni Amazon S3 o posizioni di origini dati come un database Amazon Relational Database Service (Amazon RDS). Le posizioni dei database sono sempre sedi Amazon S3.

- I servizi che si integrano con Lake Formation, come Amazon Athena e Amazon Redshift, possono accedere al Data Catalog per ottenere metadati e verificare l'autorizzazione per l'esecuzione di query. Per un elenco completo dei servizi integrati, consulta. AWSintegrazioni di servizi con Lake Formation

Argomenti

Panoramica delle autorizzazioni di Lake Formation

Esistono due tipi principali di autorizzazioni in: AWS Lake Formation

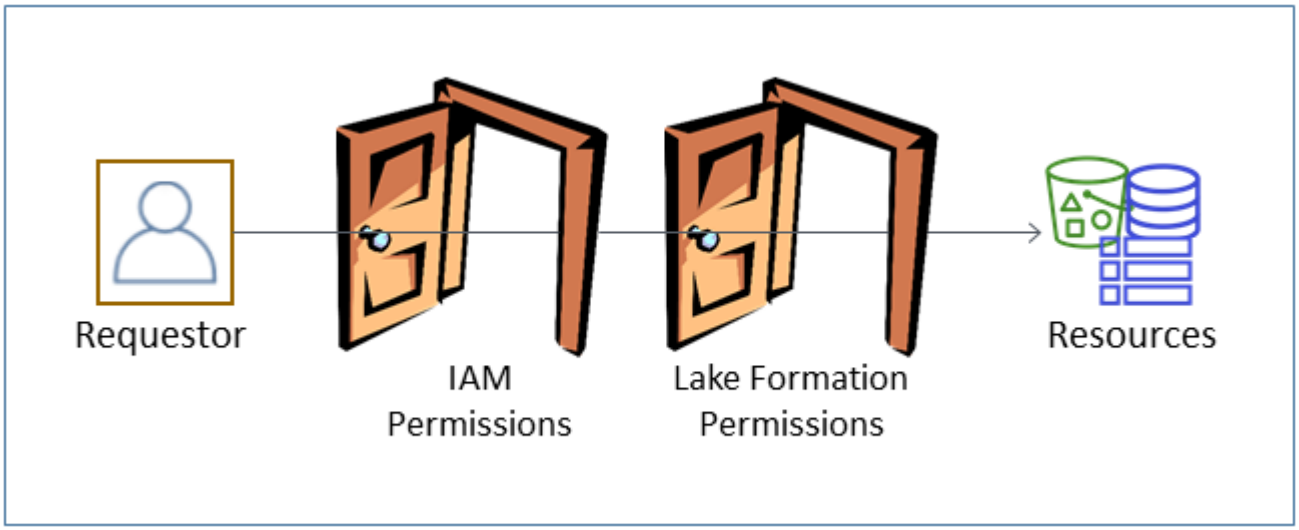
- Accesso ai metadati: autorizzazioni per le risorse del catalogo dati (autorizzazioni del catalogo dati).

  Queste autorizzazioni consentono ai responsabili di creare, leggere, aggiornare ed eliminare database e tabelle di metadati nel Catalogo dati.
- Accesso ai dati di base: autorizzazioni su sedi in Amazon Simple Storage Service (Amazon S3) (autorizzazioni di accesso ai dati e autorizzazioni di localizzazione dei dati).
  - Le autorizzazioni Data Lake consentono ai responsabili di leggere e scrivere dati nelle posizioni Amazon S3 sottostanti, dati a cui fanno riferimento le risorse del Data Catalog.

- Le autorizzazioni di localizzazione dei dati consentono ai responsabili di creare e modificare database e tabelle di metadati che puntano a posizioni Amazon S3 specifiche.

Per entrambe le aree, Lake Formation utilizza una combinazione di autorizzazioni Lake Formation e autorizzazioni AWS Identity and Access Management (IAM). Il modello di autorizzazioni IAM è costituito da politiche IAM. Il modello di permessi di Lake Formation è implementato come comandi GRANT/REVOKE in stile DBMS, come. `Grant SELECT on` *`tableName`* `to` *`userName`*

Quando un principale effettua una richiesta di accesso alle risorse del Data Catalog o ai dati sottostanti, affinché la richiesta abbia esito positivo, deve superare i controlli di autorizzazione sia di IAM che di Lake Formation.



Le autorizzazioni di Lake Formation controllano l'accesso alle risorse di Data Catalog, alle sedi Amazon S3 e ai dati sottostanti in tali sedi. Le autorizzazioni IAM controllano l'accesso a Lake Formation, alle AWS Glue API e alle risorse. Quindi, anche se potresti avere l'autorizzazione Lake Formation per creare una tabella di metadati nel Data Catalog (`CREATE_TABLE`), l'operazione fallisce se non disponi dell'autorizzazione IAM sull'`glue:CreateTable`API. (Perché un'`glue:`autorizzazione? Perché Lake Formation utilizza il AWS Glue Data Catalog.)

> **Note**
>
> Le autorizzazioni di Lake Formation si applicano solo nella regione in cui sono state concesse.

AWS Lake Formationrichiede che ogni principale (utente o ruolo) sia autorizzato a eseguire azioni sulle risorse gestite da Lake Formation. A un principale vengono concesse le autorizzazioni necessarie dall'amministratore del data lake o da un altro principale con le autorizzazioni per concedere le autorizzazioni di Lake Formation.

Quando concedi l'autorizzazione di Lake Formation a un preside, puoi facoltativamente concedere la possibilità di passare tale autorizzazione a un altro preside.

Puoi utilizzare l'API Lake Formation, il AWS Command Line Interface (AWS CLI) o le pagine Data permissions e Data locations della console Lake Formation per concedere e revocare le autorizzazioni Lake Formation.

Metodi per il controllo granulare degli accessi

Con un data lake, l'obiettivo è avere un controllo granulare degli accessi ai dati. In Lake Formation, ciò significa un controllo granulare degli accessi alle risorse del Data Catalog e alle sedi Amazon S3. Puoi ottenere un controllo granulare degli accessi con uno dei seguenti metodi.

Metodo	Autorizzazioni Lake Formation	Autorizzazioni IAM	Commenti
Metodo 1	Aperta	A grana fine	Questo è il metodo predefinito per la compatibilità con le versioni precedenti con. AWS Glue • Aperto significa che l'autorizzazione speciale `Super` viene concessa al gruppo`IAMAllowedPrincipals`, dove `IAMAllowedPrincipals` viene creata automaticamente e include tutti gli utenti e i ruoli IAM a cui è consentito l'accesso alle risorse del tuo Data Catalog dalle tue politiche IAM, e l'`Super`autorizzazione consente a un principale di eseguire ogni operazione Lake Formation supportata sul database o sulla tabella su cui è concessa. Ciò

Metodo	Autorizzazioni Lake Formation	Autorizzazioni IAM	Commenti
			fa sì che l'accesso alle risorse del Data Catalog e alle sedi Amazon S3 sia controllato esclusivamente dalle policy IAM. Per ulteriori informazioni, consultare Modifica delle impostazioni predefinite per il data lake e Aggiornamento delle autorizzazioni per AWS Glue i dati al modello AWS Lake Formation. • Con criteri granulari si intende che le policy IAM controllano tutti gli accessi alle risorse del Data Catalog e ai singoli bucket Amazon S3. Sulla console Lake Formation, questo metodo appare come Usa solo il controllo di accesso IAM.

Metodo	Autorizzazioni Lake Formation	Autorizzazioni IAM	Commenti
Metodo 2	A grana fine	A grana grossa	Questo è il metodo consigliato. • L'accesso granulare significa concedere autorizzazioni limitate di Lake Formation a singoli responsabili sulle risorse Data Catalog, sulle sedi Amazon S3 e sui dati sottostanti in tali sedi. • Graduale significa autorizzazioni più ampie per le singole operazioni e per l'accesso alle sedi Amazon S3. Ad esempio, una policy IAM a grana grossolana potrebbe includere `"glue:*"` o `"glue:Create*"` piuttosto `"glue:CreateTables"` che lasciare, le autorizzazioni di Lake Formation per controllare se un principale può creare o meno oggetti di catalogo. Significa anche dare ai responsabili l'accesso alle API di cui hanno bisogno per svolgere il loro lavoro, bloccando al contempo altre API e risorse. Ad esempio, potresti creare una policy IAM che consenta a un responsabile di creare risorse Data Catalog e creare ed eseguire flussi di lavoro, ma non consenta la creazione di AWS Glue connessioni o funzioni definite dall'utente. Vedi gli esempi più avanti in questa sezione.

> **Important**
>
> Ricorda quanto segue:
>
> - Per impostazione predefinita, Lake Formation ha le impostazioni di controllo degli accessi Use only IAM abilitate per la compatibilità con il comportamento esistente del AWS Glue Data Catalog. Ti consigliamo di disabilitare queste impostazioni dopo la transizione all'utilizzo delle autorizzazioni di Lake Formation. Per ulteriori informazioni, consulta Modifica delle impostazioni predefinite per il data lake.
> - Gli amministratori di Data Lake e i creatori di database dispongono di autorizzazioni implicite di Lake Formation che devi comprendere. Per ulteriori informazioni, consulta Autorizzazioni implicite di Lake Formation.

Controllo dell'accesso ai metadati

Per il controllo degli accessi per le risorse del Data Catalog, la discussione seguente presuppone un controllo degli accessi a grana fine con le autorizzazioni di Lake Formation e un controllo degli accessi a grana grossa con le policy IAM.

Esistono due metodi distinti per concedere le autorizzazioni di Lake Formation sulle risorse del Data Catalog:

- Controllo dell'accesso alle risorse con nome: con questo metodo, si concedono le autorizzazioni su database o tabelle specifici specificando i nomi dei database o delle tabelle. Le sovvenzioni hanno questa forma:

  Concedi le autorizzazioni ai responsabili sulle risorse [con opzione di concessione].

  Con l'opzione di concessione, puoi consentire al beneficiario di concedere le autorizzazioni ad altri mandanti.
- Controllo dell'accesso basato su tag: con questo metodo, si assegnano uno o più tag LF ai database, alle tabelle e alle colonne del Data Catalog e si concedono le autorizzazioni su uno o più tag LF ai principali. Ogni tag LF è una coppia chiave-valore, ad esempio. `department=sales` Un principale con tag LF che corrispondono ai tag LF su una risorsa del catalogo dati può accedere a tale risorsa. Questo metodo è consigliato per i data lake con un gran numero di database e tabelle. È spiegato in dettaglio inControllo degli accessi basato su tag Lake Formation.

Le autorizzazioni che un principale ha su una risorsa sono l'unione delle autorizzazioni concesse da entrambi i metodi.

La tabella seguente riassume le autorizzazioni di Lake Formation disponibili sulle risorse di Data Catalog. Le intestazioni delle colonne indicano la risorsa a cui è concessa l'autorizzazione.

Catalogo	Database	Tabella
CREATE_DATABASE	CREATE_TABLE	ALTER
	ALTER	DROP
	DROP	DESCRIBE
	DESCRIBE	SELECT*
		INSERT*
		DELETE*

Ad esempio, l'CREATE_TABLEautorizzazione viene concessa a un database. Ciò significa che il principale è autorizzato a creare tabelle in quel database.

Le autorizzazioni con un asterisco (*) sono concesse alle risorse del Data Catalog, ma si applicano ai dati sottostanti. Ad esempio, l'DROPautorizzazione per una tabella di metadati consente di eliminare la tabella dal Data Catalog. Tuttavia, l'DELETEautorizzazione concessa sulla stessa tabella consente di eliminare i dati sottostanti della tabella in Amazon S3, utilizzando, ad esempio, un'istruzione SQLDELETE. Con queste autorizzazioni, puoi anche visualizzare la tabella sulla console Lake Formation e recuperare informazioni sulla tabella con l'AWS GlueAPI. Pertanto SELECTINSERT, e DELETE sono sia le autorizzazioni di Data Catalog che le autorizzazioni di accesso ai dati.

Quando si concede la SELECT licenza su una tabella, è possibile aggiungere un filtro che includa o escluda una o più colonne. Ciò consente un controllo granulare degli accessi alle colonne della tabella di metadati, limitando le colonne che gli utenti dei servizi integrati possono vedere durante l'esecuzione delle query. Questa funzionalità non è disponibile utilizzando solo le policy IAM.

Esiste anche un'autorizzazione speciale denominataSuper. L'Superautorizzazione consente a un principale di eseguire tutte le operazioni di Lake Formation supportate sul database o sulla tabella su cui è concessa. Questa autorizzazione può coesistere con le altre autorizzazioni di Lake Formation.

Ad esempio, puoi concedere `Super` e `INSERT` su una tabella di metadati. `SELECT` Il principale può eseguire tutte le azioni supportate sulla tabella e, in caso di revoca, le autorizzazioni `SELECT` e `INSERT` rimangono `Super` invariate.

Per i dettagli su ciascuna autorizzazione, vedere. Riferimento alle autorizzazioni di Lake Formation

> **Important**
>
> Per poter vedere una tabella Data Catalog creata da un altro utente, devi avere almeno un'autorizzazione Lake Formation sulla tabella. Se ti viene concessa almeno un'autorizzazione sulla tabella, puoi anche vedere il database che contiene la tabella.

Puoi concedere o revocare le autorizzazioni di Data Catalog utilizzando la console Lake Formation, l'API o (). AWS Command Line Interface AWS CLI Di seguito è riportato un esempio di AWS CLI comando che concede all'utente il `datalake_user1` permesso di creare tabelle nel database. `retail`

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "CREATE_TABLE" --resource '{ "Database": {"Name":"retail"}}'
```

Di seguito è riportato un esempio di policy IAM di controllo degli accessi a grana grossa che integra il controllo degli accessi a grana fine con le autorizzazioni di Lake Formation. Consente tutte le operazioni su qualsiasi database o tabella di metadati.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "glue:*Database*",
                "glue:*Table*",
                "glue:*Partition*"
            ],
            "Resource": "*"
        }
    ]
}
```

L'esempio successivo è anch'esso a grana grossa ma leggermente più restrittivo. Consente operazioni di sola lettura su tutti i database e le tabelle di metadati nel Catalogo dati nell'account e nella regione designati.

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

Confronta queste politiche con la seguente politica, che implementa il controllo granulare degli accessi basato su IAM. Concede le autorizzazioni solo su un sottoinsieme di tabelle nel database di metadati di gestione delle relazioni con i clienti (CRM) nell'account e nella regione designati.

```
{
    "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*"
            ]
```

```
        }
    ]
}
```

Per altri esempi di politiche di controllo degli accessi granulari, vedere. Riferimento ai personaggi di Lake Formation e alle autorizzazioni IAM

Controllo sottostante dell'accesso ai dati

Quando un AWS servizio integrato richiede l'accesso ai dati in una posizione Amazon S3 con accesso controllato da, Lake AWS Lake Formation Formation fornisce credenziali temporanee per accedere ai dati.

Per consentire a Lake Formation di controllare l'accesso ai dati sottostanti in una sede Amazon S3, devi registrare tale posizione con Lake Formation.

Dopo aver registrato una sede Amazon S3, puoi iniziare a concedere le seguenti autorizzazioni Lake Formation:

- Autorizzazioni di accesso ai dati (`SELECT`,`INSERT`,) e `DELETE`) sulle tabelle di Data Catalog che puntano a quella posizione.
- Autorizzazioni per l'ubicazione dei dati in quella posizione.

Le autorizzazioni di localizzazione dei dati di Lake Formation controllano la capacità di creare risorse Data Catalog che puntano a particolari posizioni Amazon S3. Le autorizzazioni di localizzazione dei dati forniscono un ulteriore livello di sicurezza alle posizioni all'interno del data lake. Quando concedi l'`ALTER`autorizzazione `CREATE_TABLE` o l'autorizzazione a un'entità principale, concedi anche le autorizzazioni per la localizzazione dei dati per limitare le posizioni per le quali l'ente principale può creare o modificare tabelle di metadati.

Le sedi Amazon S3 sono bucket o prefissi all'interno di un bucket, ma non singoli oggetti Amazon S3.

Puoi concedere le autorizzazioni per la localizzazione dei dati a un principale utilizzando la console Lake Formation, l'API o ilAWS CLI. La forma generale di una sovvenzione è la seguente:

```
grant DATA_LOCATION_ACCESS to principal on S3 location [with grant option]
```

Se lo includi`with grant option`, il beneficiario può concedere le autorizzazioni ad altri mandanti.

Ricorda che le autorizzazioni di Lake Formation funzionano sempre in combinazione con le autorizzazioni AWS Identity and Access Management (IAM) per un controllo granulare degli accessi. Per le autorizzazioni di lettura/scrittura sui dati Amazon S3 sottostanti, le autorizzazioni IAM vengono concesse come segue:

Quando registri una posizione, specifichi un ruolo IAM che concede autorizzazioni di lettura/scrittura su quella posizione. Lake Formation assume questo ruolo quando fornisce credenziali temporanee ai servizi integrati. AWS Un ruolo tipico potrebbe avere la seguente politica allegata, in cui la posizione registrata è il bucket. `awsexamplebucket`

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "s3:PutObject",
                "s3:GetObject",
                "s3:DeleteObject"
            ],
            "Resource": [
                "arn:aws:s3:::awsexamplebucket/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::awsexamplebucket"
            ]
        }
    ]
}
```

Lake Formation offre un ruolo collegato al servizio che puoi utilizzare durante la registrazione per creare automaticamente politiche come questa. Per ulteriori informazioni, consulta Utilizzo di ruoli collegati ai servizi per Lake Formation.

Pertanto, la registrazione di una sede Amazon S3 concede le autorizzazioni `s3:` IAM richieste su quella posizione, dove le autorizzazioni sono specificate dal ruolo utilizzato per registrare la posizione.

> **Important**
>
> Evita di registrare un bucket Amazon S3 con Requester pay abilitato. Per i bucket registrati con Lake Formation, il ruolo utilizzato per registrare il bucket viene sempre visualizzato come richiedente. Se al bucket si accede da un altro AWS account, al proprietario del bucket viene addebitato l'accesso ai dati se il ruolo appartiene allo stesso account del proprietario del bucket.

Per l'accesso in lettura/scrittura ai dati sottostanti, oltre alle autorizzazioni di Lake Formation, i principali necessitano anche della seguente autorizzazione IAM:

`lakeformation:GetDataAccess`

Con questa autorizzazione, Lake Formation concede la richiesta di credenziali temporanee per accedere ai dati.

> **Note**
>
> Amazon Athena richiede che l'utente disponga dell'`lakeformation:GetDataAccess`autorizzazione. Altri servizi integrati richiedono che il ruolo di esecuzione sottostante disponga dell'`lakeformation:GetDataAccess`autorizzazione.

Questa autorizzazione è inclusa nelle politiche suggerite in[Riferimento ai personaggi di Lake Formation e alle autorizzazioni IAM](#).

Riassumendo, per consentire ai dirigenti di Lake Formation di leggere e scrivere i dati sottostanti con accesso controllato dai permessi di Lake Formation:

- Registra le sedi Amazon S3 che contengono i dati con Lake Formation.
- I responsabili che creano tabelle Data Catalog che puntano alle posizioni dei dati sottostanti devono disporre delle autorizzazioni per la localizzazione dei dati.

- I responsabili che leggono e scrivono i dati sottostanti devono disporre delle autorizzazioni di accesso ai dati di Lake Formation sulle tabelle del Data Catalog che puntano alle posizioni dei dati sottostanti.
- I responsabili che leggono e scrivono i dati sottostanti devono disporre dell'autorizzazione `lakeformation:GetDataAccess` IAM quando la posizione dei dati sottostanti è registrata con Lake Formation.

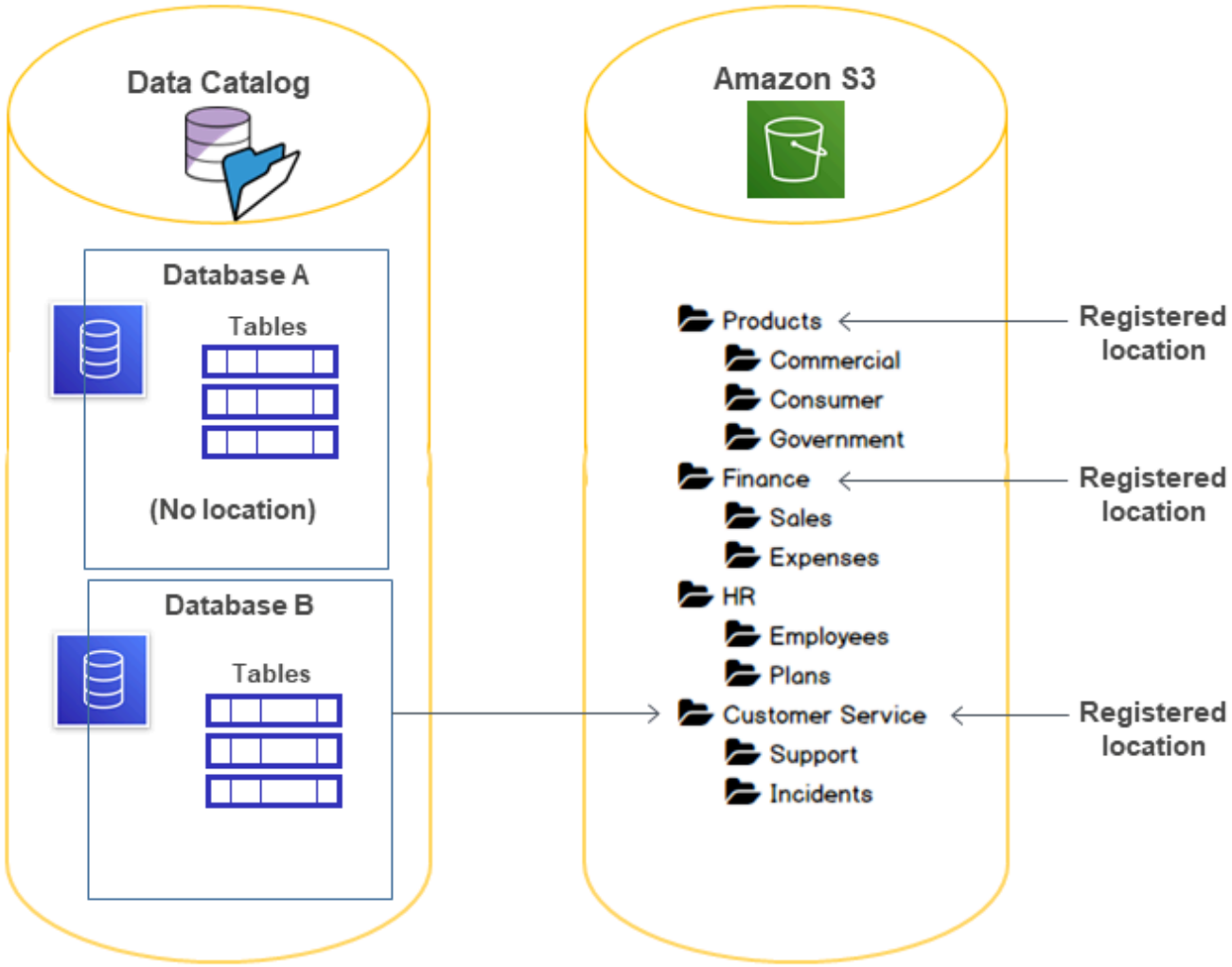
> **Note**
>
> Il modello di autorizzazioni Lake Formation non impedisce l'accesso alle sedi Amazon S3 tramite l'API o la console di Amazon S3 se hai accesso ad esse tramite policy IAM o Amazon S3. Puoi collegare le policy IAM ai principali per bloccare questo accesso.

Ulteriori informazioni sulle autorizzazioni di localizzazione dei dati

Le autorizzazioni di localizzazione dei dati regolano il risultato delle operazioni di creazione e aggiornamento su database e tabelle Data Catalog. Le regole sono le seguenti:

- Un principale deve disporre di autorizzazioni esplicite o implicite per la localizzazione dei dati su una posizione Amazon S3 per creare o aggiornare un database o una tabella che specifichi tale posizione.
- L'autorizzazione esplicita `DATA_LOCATION_ACCESS` viene concessa utilizzando la console, l'API o. AWS CLI
- Le autorizzazioni implicite vengono concesse quando un database ha una proprietà location che punta a una posizione registrata, l'entità principale dispone dell'`CREATE_TABLE`autorizzazione sul database e l'entità principale tenta di creare una tabella in quella posizione o in una posizione secondaria.
- Se a un responsabile vengono concesse le autorizzazioni per la localizzazione dei dati su una posizione, il responsabile dispone delle autorizzazioni per la localizzazione dei dati su tutte le sedi secondarie.
- Un responsabile non necessita delle autorizzazioni di localizzazione dei dati per eseguire operazioni di lettura/scrittura sui dati sottostanti. È sufficiente disporre delle `SELECT` o delle autorizzazioni di accesso ai `INSERT` dati. Le autorizzazioni per la localizzazione dei dati si applicano solo alla creazione di risorse del Catalogo Dati che puntano alla posizione.

Considerate lo scenario illustrato nel diagramma seguente.



In questo diagramma:

- I bucket `Products` Amazon S3 `Customer Service` sono registrati presso Lake Formation. `Finance`
- `Database A` non ha una proprietà di posizione e `Database B` dispone di una proprietà di posizione che punta al `Customer Service` bucket.
- L'utente `datalake_user` ha `CREATE_TABLE` su entrambi i database.
- All'utente sono `datalake_user` state concesse le autorizzazioni per la localizzazione dei dati solo nel `Products` bucket.

Di seguito sono riportati i risultati ottenuti quando l'utente `datalake_user` tenta di creare una tabella di catalogo in un particolare database in una posizione particolare.

Posizione in cui **`datalake_user`** tenta di creare una tabella

Database e posizione	Ha successo o fallisce	Motivo
Database A in `Finance/Sales`	Fallisce	Nessuna autorizzazione per la localizzazione dei dati
Database A in `Products`	Ha successo	Dispone dell'autorizzazione alla localizzazione dei dati
Database A in `HR/Plans`	Ha successo	La posizione non è registrata
Database B in `Customer Service/Incidents`	Ha successo	Il database ha la proprietà di localizzazione in `Customer Service`

Per ulteriori informazioni, consulta gli argomenti seguenti:

- Aggiungere una posizione Amazon S3 al tuo data lake
- Riferimento alle autorizzazioni di Lake Formation
- Riferimento ai personaggi di Lake Formation e alle autorizzazioni IAM

Riferimento ai personaggi di Lake Formation e alle autorizzazioni IAM

Questa sezione elenca alcuni personaggi suggeriti di Lake Formation e le relative autorizzazioni suggerite AWS Identity and Access Management (IAM). Per informazioni sulle autorizzazioni di Lake Formation, vederethe section called “Riferimento alle autorizzazioni di Lake Formation”.

AWS Lake Formation persone

La tabella seguente elenca i AWS Lake Formation personaggi suggeriti.

Personaggi di Lake Formation

Utente	Descrizione
Amministratore IAM (superutente)	(Obbligatorio) Utente che può creare utenti e ruoli IAM. Dispone della politica `AdministratorAccess` AWS gestita. Dispone di tutte le autorizzazioni su tutte le risorse di Lake Formation. Può aggiungere amministratori di data lake. Non può concedere le autorizzazioni di Lake Formation se non è stato designato anche un amministratore del data lake.
Amministratore del data lake	(Obbligatorio) Utente in grado di registrare sedi Amazon S3, accedere al Data Catalog, creare database, creare ed eseguire flussi di lavoro, concedere autorizzazioni Lake Formation ad altri utenti e visualizzare i log. AWS CloudTrail Dispone di meno autorizzazioni IAM rispetto all'amministratore IAM, ma sufficienti per amministrare il data lake. Impossibile aggiungere altri amministratori di data lake.
Amministratore di sola lettura	(Facoltativo) Utente che può visualizzare i principali, le risorse del Data Catalog, le autorizzazioni e AWS CloudTrail i registri, senza le autorizzazioni per effettuare aggiornamenti.
Ingegnere dei dati	(Facoltativo) Utente che può creare database, creare ed eseguire crawler e flussi di lavoro e concedere autorizzazioni Lake Formation sulle tabelle del Data Catalog create dai crawler e dai flussi di lavoro. Ti consigliamo di nominare tutti i data engineer creatori di database. Per ulteriori informazioni, consulta Creazione di un database.
Analista dei dati	(Facoltativo) Utente che può eseguire query sul data lake utilizzando, ad esempio,. Amazon Athena Dispone solo delle autorizzazioni sufficienti per eseguire le query.
Ruolo del workflow	(Obbligatorio) Ruolo che esegue un flusso di lavoro per conto di un utente. Questo ruolo viene specificato quando si crea un flusso di lavoro da un blueprint.

AWS politiche gestite per Lake Formation

Puoi concedere le autorizzazioni AWS Identity and Access Management (IAM) necessarie per lavorare AWS Lake Formation utilizzando policy AWS gestite e policy in linea. Le seguenti politiche AWS gestite sono disponibili per Lake Formation.

AWS politica gestita: AWSLakeFormationDataAdmin

AWSLakeFormationDataAdminla politica garantisce l'accesso amministrativo AWS Lake Formation ai servizi correlati, come AWS Glue la gestione dei data lake.

Puoi collegarti `AWSLakeFormationDataAdmin` ai tuoi utenti, gruppi e ruoli.

Dettagli sulle autorizzazioni

- `CloudTrail`— Consente ai responsabili di visualizzare i AWS CloudTrail registri. Ciò è necessario per esaminare eventuali errori nella configurazione del data lake.
- `Glue`— Consente ai responsabili di visualizzare, creare e aggiornare tabelle e database di metadati in Data Catalog. Ciò include le operazioni API che iniziano con`Get`,`List`,, `Create` `UpdateDelete`, e. `Search` Ciò è necessario per gestire i metadati delle tabelle del data lake.
- `IAM`— Consente ai responsabili di recuperare informazioni sugli utenti, i ruoli e le policy IAM associati ai ruoli. Ciò è necessario affinché l'amministratore dei dati riveda ed elenchi gli utenti e i ruoli IAM a cui concedere le autorizzazioni di Lake Formation.
- `Lake Formation`— Concede agli amministratori dei data lake le autorizzazioni necessarie per Lake Formation per gestire i data lake.
- `S3`— Consente ai responsabili di recuperare informazioni sui bucket Amazon S3 e sulle loro posizioni per configurare la posizione dei dati per i data lake.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:*",
                "cloudtrail:DescribeTrails",
                "cloudtrail:LookupEvents",
                "glue:GetDatabase",
                "glue:GetDatabases",
```

```
                "glue:CreateDatabase",
                "glue:UpdateDatabase",
                "glue:DeleteDatabase",
                "glue:GetConnections",
                "glue:SearchTables",
                "glue:GetTable",
                "glue:CreateTable",
                "glue:UpdateTable",
                "glue:DeleteTable",
                "glue:GetTableVersions",
                "glue:GetPartitions",
                "glue:GetTables",
                "glue:GetWorkflow",
                "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": "*"
        },
        {
            "Effect": "Deny",
            "Action": [
                "lakeformation:PutDataLakeSettings"
            ],
            "Resource": "*"
        }
    ]
}
```

> **Note**
>
> La `AWSLakeFormationDataAdmin` policy non concede tutte le autorizzazioni necessarie agli amministratori dei data lake. Sono necessarie autorizzazioni aggiuntive per creare ed eseguire flussi di lavoro e registrare posizioni con il ruolo collegato al servizio. `AWSServiceRoleForLakeFormationDataAccess` Per ulteriori informazioni, consultare Crea un amministratore del data lake e Utilizzo di ruoli collegati ai servizi per Lake Formation.

AWS politica gestita: AWSLakeFormationCrossAccountManager

AWSLakeFormationCrossAccountManagerla politica fornisce l'accesso multiaccount alle AWS Glue risorse tramite Lake Formation e concede l'accesso in lettura ad altri servizi richiesti come AWS Organizations e AWS RAM.

Puoi collegarti `AWSLakeFormationCrossAccountManager` ai tuoi utenti, gruppi e ruoli.

Dettagli sulle autorizzazioni

Questa policy include le seguenti autorizzazioni:

- `Glue`— Consente ai responsabili di impostare o eliminare la politica delle risorse del Data Catalog per il controllo degli accessi.
- `Organizations`— Consente ai responsabili di recuperare le informazioni sull'account e sull'unità organizzativa (OU) di un'organizzazione.
- `ram:CreateResourceShare`— Consente ai dirigenti di creare una condivisione di risorse.
- `ram:UpdateResourceShare`— Consente ai principali di modificare alcune proprietà della condivisione di risorse specificata.
- `ram:DeleteResourceShare`— Consente ai principali di eliminare la condivisione di risorse specificata.
- `ram:AssociateResourceShare`— Consente ai responsabili di aggiungere l'elenco di principali e l'elenco di risorse specificati a una condivisione di risorse.
- `ram:DisassociateResourceShare`— Consente ai principali di rimuovere i principali o le risorse specificati dalla partecipazione alla condivisione di risorse specificata.
- `ram:GetResourceShares`— Consente ai responsabili di recuperare i dettagli sulle condivisioni di risorse che possiedi o che sono condivise con te.

- `ram:RequestedResourceType`— Consente ai responsabili di recuperare il tipo di risorsa (database, tabella o catalogo).
- `AssociateResourceSharePermission`— Consente ai responsabili di aggiungere o sostituire l' AWS RAM autorizzazione per un tipo di risorsa incluso in una condivisione di risorse. È possibile associare esattamente un'autorizzazione a ciascun tipo di risorsa nella condivisione di risorse.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ram:CreateResourceShare"
            ],
            "Resource": "*",
            "Condition": {
                "StringLikeIfExists": {
                    "ram:RequestedResourceType": [
                        "glue:Table",
                        "glue:Database",
                        "glue:Catalog"
                    ]
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": [
                "ram:UpdateResourceShare",
                "ram:DeleteResourceShare",
                "ram:AssociateResourceShare",
                "ram:DisassociateResourceShare",
                "ram:GetResourceShares"
            ],
            "Resource": "*",
            "Condition": {
                "StringLike": {
                    "ram:ResourceShareName": [
                        "LakeFormation*"
                    ]
                }
```

```
            }
        },
        {
            "Effect": "Allow",
            "Action": [
                "ram:AssociateResourceSharePermission"
            ],
            "Resource": "*",
            "Condition": {
                "StringLike": {
                    "ram:PermissionArn": [
                        "arn:aws:ram::aws:permission/AWSRAMLFEnabled*"
                    ]
                }
            }
       },
        {
            "Effect": "Allow",
            "Action": [
                "glue:PutResourcePolicy",
                "glue:DeleteResourcePolicy",
                "organizations:DescribeOrganization",
                "organizations:DescribeAccount",
                "ram:Get*",
                "ram:List*"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "organizations:ListRoots",
                "organizations:ListAccountsForParent",
                "organizations:ListOrganizationalUnitsForParent"
            ],
            "Resource": "*"
        }
    ]
}
```

AWS politica gestita: AWSGlueConsoleFullAccess

AWSGlueConsoleFullAccessla politica garantisce l'accesso completo alle AWS Glue risorse quando un'identità a cui è associata la politica utilizza il AWS Management Console. Se segui la convenzione per la denominazione per le risorse specificate nella policy, gli utenti hanno la piena funzionalità della console. Questo criterio è in genere associato agli utenti della AWS Glue console.

Inoltre, AWS Glue Lake Formation assume il ruolo di servizio `AWSGlueServiceRole` per consentire l'accesso ai servizi correlati, tra cui Amazon Elastic Compute Cloud (Amazon EC2), Amazon Simple Storage Service (Amazon S3) e Amazon. CloudWatch

AWS managed policy:LakeFormationDataAccessServiceRolePolicy

Questa policy è associata a un ruolo collegato al servizio denominato `ServiceRoleForLakeFormationDataAccess` che consente al servizio di eseguire azioni sulle risorse su richiesta dell'utente. Non puoi collegare questa policy alle tue identità IAM.

Questa policy consente ai AWS servizi integrati di Lake Formation come Amazon Athena Amazon Redshift di utilizzare il ruolo collegato ai servizi per scoprire le risorse di Amazon S3.

Per ulteriori informazioni, consulta Utilizzo di ruoli collegati ai servizi per Lake Formation.

Dettagli di autorizzazione

Questa politica include le seguenti autorizzazioni.

- `s3:ListAllMyBuckets`— Restituisce un elenco di tutti i bucket di proprietà del mittente autenticato della richiesta.

```
{
 "Version": "2012-10-17",
 "Statement": [
  {
   "Sid": "LakeFormationDataAccessServiceRolePolicy",
   "Effect": "Allow",
   "Action": [
    "s3:ListAllMyBuckets"
   ],
   "Resource": [
    "arn:aws:s3:::*"
   ]
```

```
  }
 ]
}
```

Lake Formation: aggiornamenti alle politiche AWS gestite

Visualizza i dettagli sugli aggiornamenti alle politiche AWS gestite per Lake Formation da quando questo servizio ha iniziato a tracciare queste modifiche.

Modifica	Descrizione	Data
`LakeFormationDataAccessServiceRolePolicy` Politica aggiornata di Lake Formation.	Lake Formation ha migliorato la LakeFormationDataAccessServiceRolePolicypolitica aggiungendo un elemento Sid alla dichiarazione politica.	febbraio 2024
`AWSLakeFormationCrossAccountManager` Politica aggiornata di Lake Formation.	Lake Formation ha migliorato la AWSLakeFormationCrossAccountManagerpolitica aggiungendo una nuova autorizzazione per abilitare la condivisione dei dati tra account in modalità di accesso ibrido.	ottobre 2023
`AWSLakeFormationCrossAccountManager` Politica aggiornata di Lake Formation.	Lake Formation ha migliorato la AWSLakeFormationCrossAccountManagerpolitica per creare una sola condivisione di risorse per account destinatario quando la risorsa viene condivisa per la prima volta. Tutte le risorse condivise successivamente con lo stesso account vengono allegate alla stessa condivisione di risorse.	6 maggio 2022
Lake Formation ha iniziato a tracciare le modifiche.	Lake Formation ha iniziato a tracciare le modifiche per le sue politiche AWS gestite.	6 maggio 2022

Autorizzazioni suggerite da Personas

Di seguito sono riportate le autorizzazioni suggerite per ogni persona. L'amministratore IAM non è incluso perché quell'utente dispone di tutte le autorizzazioni su tutte le risorse.

Argomenti

Autorizzazioni di amministratore di Data Lake

> **Important**
>
> Nei seguenti criteri, sostituiscili <account-id>con un numero di AWS account valido e sostituiscili <workflow_role>con il nome di un ruolo che dispone delle autorizzazioni per eseguire un flusso di lavoro, come definito in. Autorizzazioni relative ai ruoli del workflow

Tipo di policy	Policy
AWS politiche gestite	• `AWSLakeFormationDataAdmin` • `LakeFormationDataAccessServiceRolePolicy` (politica relativa ai ruoli legati ai servizi) • `AWSGlueConsoleFullAccess` (facoltativo). • `CloudWatchLogsReadOnlyAccess` (facoltativo) • `AWSLakeFormationCrossAccountManager` (facoltativo) • `AmazonAthenaFullAccess` (facoltativo). Per informazioni sulle politiche AWS gestite opzionali, vedere. the section called “Crea un amministratore del data lake”

Tipo di policy	Policy

Politica in linea (per la creazione del ruolo collegato ai servizi di 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 di policy	Policy

(Facoltativo) Politica in linea (politica passrole per il ruolo del flusso di lavoro). Questa operazione è necessaria solo se l'amministratore del data lake crea ed esegue flussi di lavoro.	<pre>{
	    "Version": "2012-10-17",
	    "Statement": [
	        {
	            "Sid": "PassRolePermissions",
	            "Effect": "Allow",
	            "Action": [
	                "iam:PassRole"
	            ],
	            "Resource": [
	                "arn:aws:iam:: <account-id> :role/<workflow_role> "
	            ]
	        }
	    ]
	}</pre>

(Facoltativo) Politica in linea (se il tuo account concede o riceve autorizzazioni Lake Formation per più account). Questa politica serve ad accettare o rifiutare gli inviti alla condivisione di AWS RAM risorse e a consentire la concessione di autorizzazioni per più account alle organizzazioni. `ram:EnableSharingWithAwsOrganization` è richiesto solo per gli amministratori del data lake presenti nell'account di gestione. AWS Organizations	<pre>{
	    "Version": "2012-10-17",
	    "Statement": [
	        {
	            "Effect": "Allow",
	            "Action": [
	                "ram:AcceptResourceShareInvitation",
	                "ram:RejectResourceShareInvitation",
	                "ec2:DescribeAvailabilityZones",
	                "ram:EnableSharingWithAwsOrganization"
	            ],
	            "Resource": "*"
	        }
	    ]
	}</pre>

Autorizzazioni di amministratore di sola lettura

Tipo di policy	Policy
Politica in linea (di base)	{ "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:ListLakeFormationOption", "cloudtrail:DescribeTrails", "cloudtrail:LookupEvents", "glue:GetDatabase", "glue:GetDatabases", "glue:GetConnections", "glue:SearchTables", "glue:GetTable", "glue:GetTableVersions", "glue:GetPartitions", "glue:GetTables", "glue:GetWorkflow", "glue:ListWorkflows", "glue:BatchGetWorkflows", "glue:GetWorkflowRuns", "glue:GetWorkflow", "s3:ListBucket", "s3:GetBucketLocation", "s3:ListAllMyBuckets", "s3:GetBucketAcl", "iam:ListUsers",

Tipo di policy	Policy

	"iam:ListRoles",
	"iam:GetRole",
	"iam:GetRolePolicy"
	],
	"Resource":"*"
	},
	{
	"Effect":"Deny",
	"Action":[
	"lakeformation:PutDataLakeSettings"
	],
	"Resource":"*"
	}
	]
	}

Autorizzazioni del tecnico dei dati

> **Important**
>
> Nelle seguenti politiche, sostituiscilo <account-id>con un numero di AWS account valido e sostituiscilo <workflow_role>con il nome del ruolo del flusso di lavoro.

Tipo di policy	Policy
AWS politica gestita	`AWSGlueConsoleFullAccess`
politica in linea (di base)	{ "Version": "2012-10-17", "Statement": [{ "Effect": "Allow", "Action": ["lakeformation:GetDataAccess", "lakeformation:GrantPermissions", "lakeformation:RevokePermissions", "lakeformation:BatchGrantPermissions",

Tipo di policy	Policy
	"lakeformation:BatchRevokePermissions", "lakeformation:ListPermissions", "lakeformation:AddLFTagsToResource", "lakeformation:RemoveLFTagsFromResou rce", "lakeformation:GetResourceLFTags", "lakeformation:ListLFTags", "lakeformation:GetLFTag", "lakeformation:SearchTablesByLFTags", "lakeformation:SearchDatabasesByLFTa gs", "lakeformation:GetWorkUnits", "lakeformation:GetWorkUnitResults", "lakeformation:StartQueryPlanning", "lakeformation:GetQueryState", "lakeformation:GetQueryStatistics"], "Resource": "*" }] }

Tipo di policy	Policy

Politica in linea (per le operazioni su tabelle gestite, incluse le operazioni all'interno delle transazioni)	{
	"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 di policy	Policy

Policy in linea (per il controllo dell'accesso ai metadati utilizzando il metodo di controllo degli accessi basato su tag Lake Formation (LF-TBAC))	{
	"Version": "2012-10-17",
	"Statement": [
	{
	"Effect": "Allow",
	"Action": [
	"lakeformation:AddLFTagsToResource",
	"lakeformation:RemoveLFTagsFromResou
	rce",
	"lakeformation:GetResourceLFTags",
	"lakeformation:ListLFTags",
	"lakeformation:GetLFTag",
	"lakeformation:SearchTablesByLFTags",
	"lakeformation:SearchDatabasesByLFTags"
	],
	"Resource": "*"
	}
	]
	}

Politica in linea (politica passrole per il ruolo del flusso di lavoro)	{
	"Version": "2012-10-17",
	"Statement": [
	{
	"Sid": "PassRolePermissions",
	"Effect": "Allow",
	"Action": [
	"iam:PassRole"
	],
	"Resource": [
	"arn:aws:iam:: *<account-id>* :role/*<workflow
	_role>* "
	]
	}
	]
	}

Autorizzazioni per analisti di dati

Tipo di policy	Policy
AWS politica gestita	`AmazonAthenaFullAccess`
politica in linea (di base)	`{ "Version": "2012-10-17", "Statement": [{ "Effect": "Allow", "Action": ["lakeformation:GetDataAccess", "glue:GetTable", "glue:GetTables", "glue:SearchTables", "glue:GetDatabase", "glue:GetDatabases", "glue:GetPartitions", "lakeformation:GetResourceLFTags", "lakeformation:ListLFTags", "lakeformation:GetLFTag", "lakeformation:SearchTablesByLFTags", "lakeformation:SearchDatabasesByLFTags"], "Resource": "*" }] }`
(Facoltativo) Politica in linea (per le operazioni su tabelle gestite, incluse le operazioni all'interno delle transazioni)	`{ "Version": "2012-10-17", "Statement": [ { "Effect": "Allow", "Action": [ "lakeformation:StartTransaction", "lakeformation:CommitTransaction", "lakeformation:CancelTransaction", "lakeformation:ExtendTransaction", "lakeformation:DescribeTransaction",`

Tipo di policy	Policy
	"lakeformation:ListTransactions", "lakeformation:GetTableObjects", "lakeformation:UpdateTableObjects", "lakeformation:DeleteObjectsOnCancel"], "Resource": "*" }] }

Autorizzazioni relative ai ruoli del workflow

Questo ruolo dispone delle autorizzazioni necessarie per eseguire un flusso di lavoro. Si specifica un ruolo con queste autorizzazioni quando si crea un flusso di lavoro.

> **Important**
>
> Nelle seguenti politiche, sostituiscilo <region>con un identificatore di AWS regione valido (ad esempious-east-1), <account-id>con un numero di AWS account valido, <workflow_role>con il nome del ruolo del flusso di lavoro e <your-s3-cloudtrail-bucket>con il percorso Amazon S3 verso AWS CloudTrail i log.

Tipo di policy	Policy
AWS politica gestita	`AWSGlueServiceRole`
policy in linea (accesso ai dati)	{ "Version": "2012-10-17", "Statement": [{ "Sid": "Lakeformation", "Effect": "Allow", "Action": ["lakeformation:GetDataAccess", "lakeformation:GrantPermissions"],

Tipo di policy	Policy

	<pre>"Resource": "*"
	}
	]
	}</pre>

Politica in linea (politica passrole per il ruolo del flusso di lavoro)	<pre>{
	"Version": "2012-10-17",
	"Statement": [
	{
	"Sid": "PassRolePermissions",
	"Effect": "Allow",
	"Action": [
	"iam:PassRole"
	],
	"Resource": [
	"arn:aws:iam:: <account-id> :role/<workflow_role> "
	]
	}
	]
	}</pre>

Policy in linea (per l'acquisizione di dati all'esterno del data lake, ad esempio log) AWS CloudTrail	<pre>{
	"Version": "2012-10-17",
	"Statement": [
	{
	"Effect": "Allow",
	"Action": ["s3:GetObject", "s3:ListBucket"],
	"Resource": ["arn:aws:s3::: <your-s3-cloudtrail-bucket> /*"]
	}
	]
	}</pre>

Modifica delle impostazioni predefinite per il data lake

Per mantenere la retrocompatibilità conAWS Glue, AWS Lake Formation dispone delle seguenti impostazioni di sicurezza iniziali:

- L'`Super`autorizzazione viene concessa al gruppo `IAMAllowedPrincipals` su tutte le risorse del AWS Glue Data Catalog esistenti.
- Le impostazioni «Usa solo il controllo di accesso IAM» sono abilitate per le nuove risorse del Data Catalog.

Queste impostazioni fanno sì che l'accesso alle risorse di Data Catalog e alle sedi Amazon S3 sia controllato esclusivamente da policy AWS Identity and Access Management (IAM). Le autorizzazioni individuali di Lake Formation non sono valide.

Il `IAMAllowedPrincipals` gruppo include tutti gli utenti e i ruoli IAM a cui è consentito l'accesso alle risorse del Data Catalog in base alle politiche IAM. L'`Super`autorizzazione consente a un principale di eseguire tutte le operazioni di Lake Formation supportate sul database o sulla tabella su cui è concessa.

Per modificare le impostazioni di sicurezza in modo che l'accesso alle risorse del Data Catalog (database e tabelle) sia gestito dalle autorizzazioni di Lake Formation, procedi come segue:

1. Modifica le impostazioni di sicurezza predefinite per le nuove risorse. Per istruzioni, consulta Modifica il modello di autorizzazione predefinito o utilizza la modalità di accesso ibrida.
2. Modifica le impostazioni per le risorse esistenti del Data Catalog. Per istruzioni, consulta Aggiornamento delle autorizzazioni per AWS Glue i dati al modello AWS Lake Formation.

Modifica delle impostazioni di sicurezza predefinite utilizzando l'operazione dell'**`PutDataLakeSettings`**API Lake Formation

Puoi anche modificare le impostazioni di sicurezza predefinite utilizzando l'operazione PutDataLakeSettingsAPI Lake Formation. Questa azione utilizza come argomenti un ID di catalogo e una DataLakeSettingsstruttura opzionali.

Per applicare i metadati e il controllo dell'accesso ai dati sottostanti da parte di Lake Formation su nuovi database e tabelle, codifica la `DataLakeSettings` struttura come segue.

> **Note**
>
> Sostituisci <AccountID>con un ID AWS account valido e <Username>con un nome utente IAM valido. Puoi specificare più di un utente come amministratore del data lake.

```
{
    "DataLakeSettings": {
        "DataLakeAdmins": [
            {
                "DataLakePrincipalIdentifier":
 "arn:aws:iam::<AccountId>:user/<Username>"
            }
        ],
        "CreateDatabaseDefaultPermissions": [],
        "CreateTableDefaultPermissions": []
    }
}
```

È inoltre possibile codificare la struttura come segue. Omettere il `CreateTableDefaultPermissions` parametro `CreateDatabaseDefaultPermissions` or equivale a passare un elenco vuoto.

```
{
    "DataLakeSettings": {
        "DataLakeAdmins": [
            {
                "DataLakePrincipalIdentifier":
 "arn:aws:iam::<AccountId>:user/<Username>"
            }
        ]
    }
}
```

Questa azione revoca efficacemente tutte le autorizzazioni di Lake Formation dal `IAMAllowedPrincipals` gruppo su nuovi database e tabelle. Quando crei un database, puoi sovrascrivere questa impostazione.

Per applicare i metadati e il controllo dell'accesso ai dati sottostanti solo da parte di IAM su nuovi database e tabelle, codifica la `DataLakeSettings` struttura come segue.

```
{
    "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"
                ]
            }
        ]
    }
}
```

Ciò concede a `Super` Lake Formation l'autorizzazione al `IAMAllowedPrincipals` gruppo su nuovi database e tabelle. Quando crei un database, puoi sovrascrivere questa impostazione.

> **Note**
>
> Nella `DataLakeSettings` struttura precedente, l'unico valore consentito per `DataLakePrincipalIdentifier` è e `IAM_ALLOWED_PRINCIPALS` l'unico valore consentito per `Permissions` è. `ALL`

Autorizzazioni implicite di Lake Formation

AWS Lake Formationconcede le seguenti autorizzazioni implicite agli amministratori di data lake, ai creatori di database e ai creatori di tabelle.

amministratori di data lake

- Hanno `Describe` accesso a tutte le risorse del Data Catalog ad eccezione delle risorse condivise da un altro account direttamente con un altro principale. Questo accesso non può essere revocato a un amministratore.
- Disponi delle autorizzazioni per la localizzazione dei dati ovunque nel data lake.
- Può concedere o revocare l'accesso a qualsiasi risorsa del Data Catalog a qualsiasi principale (incluso se stesso). Questo accesso non può essere revocato a un amministratore.
- Può creare database nel Data Catalog.
- Può concedere il permesso di creare un database a un altro utente.

> **Note**
>
> Gli amministratori di Data Lake possono registrare le sedi Amazon S3 solo se dispongono delle autorizzazioni IAM necessarie per farlo. Le politiche di amministrazione dei data lake suggerite in questa guida concedono tali autorizzazioni. Inoltre, gli amministratori dei data lake non dispongono delle autorizzazioni implicite per eliminare database o alterare/eliminare tabelle create da altri. Tuttavia, possono concedersi le autorizzazioni per farlo.

Per ulteriori informazioni sugli amministratori del data lake, consulta. Crea un amministratore del data lake

Creatori di database

- Dispongono di tutte le autorizzazioni di database sui database che creano, delle autorizzazioni sulle tabelle che creano nel database e possono concedere ad altri responsabili dello stesso AWS account il permesso di creare tabelle nel database. Un creatore di database che dispone anche della politica `AWSLakeFormationCrossAccountManager` AWS gestita può concedere le autorizzazioni sul database ad altri AWS account o organizzazioni.

  Gli amministratori di Data Lake possono utilizzare la console o l'API di Lake Formation per designare i creatori di database.

  > **Note**
  >
  > I creatori di database non dispongono implicitamente delle autorizzazioni sulle tabelle create da altri nel database.

Per ulteriori informazioni, consulta Creazione di un database.

Creatori di tabelle

- Disponi di tutte le autorizzazioni sulle tabelle che creano.
- Possono concedere le autorizzazioni su tutte le tabelle che creano ai responsabili dello stesso account. AWS
- Può concedere le autorizzazioni su tutte le tabelle che crea ad altri AWS account o organizzazioni se dispongono della `AWSLakeFormationCrossAccountManager` AWS politica gestita.
- Può visualizzare i database che contengono le tabelle che crea.

Riferimento alle autorizzazioni di Lake Formation

Per eseguire AWS Lake Formation le operazioni, i mandanti necessitano sia delle autorizzazioni di Lake Formation che delle autorizzazioni AWS Identity and Access Management (IAM). In genere concedi le autorizzazioni IAM utilizzando politiche di controllo degli accessi granulari, come descritto in. the section called “Panoramica delle autorizzazioni di Lake Formation ” Puoi concedere le autorizzazioni di Lake Formation utilizzando la console, l'API o AWS Command Line Interface (AWS CLI).

Per informazioni su come concedere o revocare le autorizzazioni di Lake Formation, vedere e. the section called “Concessione e revoca delle autorizzazioni di Data Catalog” the section called “Concessione delle autorizzazioni per la localizzazione dei dati”

> **Note**
>
> Gli esempi in questa sezione mostrano come concedere le autorizzazioni ai responsabili dello stesso account. AWS Per esempi di sovvenzioni su più account, vedere. the section called “Condivisione dei dati tra account”

Autorizzazioni Lake Formation per tipo di risorsa

Di seguito sono riportate le autorizzazioni valide di Lake Formation disponibili per ogni tipo di risorsa:

Risorsa	Autorizzazione
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
Amazon S3 location	DATA_LOCATION_ACCESS
LF-Tags	DROP
	ALTER

Risorsa	Autorizzazione
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
Table with data filters	DESCRIBE
	DROP
	SELECT

Risorsa	Autorizzazione
`Table with column filter`	`SELECT`

Argomenti

Comandi di concessione e AWS CLI revoca di Lake Formation

Ogni descrizione dell'autorizzazione in questa sezione include esempi di concessione dell'autorizzazione tramite un comando. AWS CLI Di seguito sono riportate le sinossi grant-permissions e revoke-permissions AWS CLI i comandi del Lake Formation.

```
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>]
```

Per una descrizione dettagliata di questi comandi, vedete grant-permissions e revoke-permissions nel Command Reference.AWS CLI Questa sezione fornisce informazioni aggiuntive sull'opzione. `--principal`

Il valore dell'`--principal`opzione è uno dei seguenti:

- Amazon Resource Name (ARN) per un utente o un AWS Identity and Access Management ruolo (IAM)
- ARN per un utente o un gruppo che esegue l'autenticazione tramite un provider SAML, come Microsoft Active Directory Federation Service (ADFS)
- ARN per un QuickSight utente o un gruppo Amazon
- Per le autorizzazioni su più account, un ID AWS account, un ID dell'organizzazione o un ID di unità organizzativa

Di seguito sono riportati la sintassi e gli esempi per tutti i tipi. `--principal`

Principal è un utente IAM

Sintassi:

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/<user-name>
```

Esempio:

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1
```

Principal è un ruolo IAM

Sintassi:

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:role/<role-name>
```

Esempio:

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:role/workflowrole
```

Principal è un utente che si autentica tramite un provider SAML

Sintassi:

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:saml-
provider/<SAMLproviderName>:user/<user-name>
```

Esempi:

```
--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
```

Principal è un gruppo che si autentica tramite un provider SAML

Sintassi:

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:saml-
provider/<SAMLproviderName>:group/<group-name>
```

Esempi:

```
--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 è un utente di Amazon QuickSight Enterprise Edition

Sintassi:

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:<region>:<account-
id>:user/<namespace>/<user-name>
```

> **Note**
>
> Perché<namespace>, devi specificare`default`.

Esempio:

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:us-
east-1:111122223333:user/default/bi_user1
```

Principal è un gruppo Amazon QuickSight Enterprise Edition

Sintassi:

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:<region>:<account-id>:group/<namespace>/<group-name>
```

> **Note**
>
> Perché<namespace>, devi specificaredefault.

Esempio:

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:us-east-1:111122223333:group/default/data_scientists
```

Il principale è un AWS account

Sintassi:

```
--principal DataLakePrincipalIdentifier=<account-id>
```

Esempio:

```
--principal DataLakePrincipalIdentifier=111122223333
```

Il preside è un'organizzazione

Sintassi:

```
--principal DataLakePrincipalIdentifier=arn:aws:organizations::<account-id>:organization/<organization-id>
```

Esempio:

```
--principal
 DataLakePrincipalIdentifier=arn:aws:organizations::111122223333:organization/o-abcdefghijkl
```

Il preside è un'unità organizzativa

Sintassi:

```
--principal DataLakePrincipalIdentifier=arn:aws:organizations::<account-id>:ou/<organization-id>/<organizational-unit-id>
```

Esempio:

```
--principal DataLakePrincipalIdentifier=arn:aws:organizations::111122223333:ou/o-abcdefghijkl/ou-ab00-cdefghij
```

Il preside è un utente o un gruppo di identità di IAM Identity Center

Esempio: utente

```
--principal DataLakePrincipalIdentifier=arn:aws:identitystore:::user/<UserID>
```

Esempio: Gruppo:

```
--principal DataLakePrincipalIdentifier=arn:aws:identitystore:::group/<GroupID>
```

Principal è un gruppo IAM - **`IAMAllowedPrincipals`**

Lake Formation imposta le `Super` autorizzazioni su tutti i database e le tabelle del Data Catalog a un gruppo chiamato per `IAMAllowedPrincipals` impostazione predefinita. Se questa autorizzazione di gruppo esiste su un database o una tabella, tutti i principali del tuo account avranno accesso alla risorsa tramite le politiche principali IAM per AWS Glue. Fornisce la compatibilità con le versioni precedenti quando inizi a utilizzare le autorizzazioni di Lake Formation per proteggere le risorse del Data Catalog per cui in precedenza erano protette dalle politiche IAM. AWS Glue

Quando utilizzi Lake Formation per gestire le autorizzazioni per le tue risorse Data Catalog, devi prima revocare l'`IAMAllowedPrincipals`autorizzazione sulle risorse o attivare i principali e le risorse per la modalità di accesso ibrido affinché le autorizzazioni di Lake Formation funzionino.

Esempio:

```
--principal DataLakePrincipalIdentifier=IAM_Allowed_Principals
```

Principal è un gruppo IAM - **`ALLIAMPrincipals`**

Quando concedi le autorizzazioni per il `ALLIAMPrincipals` gruppo su una risorsa Data Catalog, ogni principale dell'account ottiene l'accesso alla risorsa Data Catalog utilizzando le autorizzazioni Lake Formation e le autorizzazioni IAM.

Esempio:

```
--principal DataLakePrincipalIdentifier=123456789012:IAMPrincipals
```

Autorizzazioni Lake Formation

Questa sezione contiene i permessi disponibili di Lake Formation che puoi concedere ai presidi.

`ALTER`

Autorizzazione	Concesso su questa risorsa	Inoltre, il beneficiario ha bisogno
`ALTER`	`DATABASE`	`glue:UpdateDatabase`
`ALTER`	`TABLE`	`glue:UpdateTable`
`ALTER`	`LF-Tag`	`lakeformation:UpdateLFTag`

Un principale con questa autorizzazione può modificare i metadati di un database o di una tabella nel Catalogo dati. Per le tabelle, è possibile modificare lo schema delle colonne e aggiungere parametri di colonna. Non è possibile modificare le colonne nei dati sottostanti a cui fa riferimento una tabella di metadati.

Se la proprietà da modificare è una posizione registrata di Amazon Simple Storage Service (Amazon S3), il principale deve disporre delle autorizzazioni per la localizzazione dei dati nella nuova posizione.

Example

L'esempio seguente concede l'`ALTER`autorizzazione all'utente `datalake_user1` sul database `retail` nell'account 1111-2222-3333. AWS

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "ALTER" --resource '{ "Database": {"Name":"retail"}}'
```

Example

L'esempio seguente concede `ALTER` a un utente l'`datalake_user1`accesso alla tabella del database. `inventory` `retail`

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "ALTER" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

CREATE_DATABASE

Autorizzazione	Concesso su questa risorsa	Inoltre, il beneficiario ha bisogno
`CREATE_DATABASE`	Catalogo dati	`glue:CreateDatabase`

Un responsabile con questa autorizzazione può creare un database di metadati o un collegamento a una risorsa nel Catalogo dati. Il principale può anche creare tabelle nel database.

Example

L'esempio seguente concede `CREATE_DATABASE` a un utente `datalake_user1` nell' AWS account 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "CREATE_DATABASE" --resource '{ "Catalog": {}}'
```

Quando un principale crea un database nel Data Catalog, non viene concessa alcuna autorizzazione per i dati sottostanti. Vengono concesse le seguenti autorizzazioni aggiuntive per i metadati (oltre alla possibilità di concedere tali autorizzazioni ad altri):

- `CREATE_TABLE`nel database

- Database `ALTER`
- Database `DROP`

Quando crea un database, il principale può facoltativamente specificare una posizione Amazon S3. A seconda che il principale disponga delle autorizzazioni per la localizzazione dei dati, l'`CREATE_DATABASE`autorizzazione potrebbe non essere sufficiente per creare database in tutti i casi. È importante tenere a mente i tre casi seguenti.

Crea un caso d'uso del database	Autorizzazioni necessarie
La proprietà location non è specificata.	`CREATE_DATABASE` è sufficiente.
La proprietà location è specificata e la posizione non è gestita da Lake Formation (non è registrata).	`CREATE_DATABASE` è sufficiente.
La proprietà location è specificata e la posizione è gestita da Lake Formation (è registrata).	`CREATE_DATABASE` è obbligatorio più le autorizzazioni di localizzazione dei dati nella posizione specificata.

`CREATE_TABLE`

Autorizzazione	Concesso su questa risorsa	Inoltre, il beneficiario ha bisogno
`CREATE_TABLE`	`DATABASE`	`glue:CreateTable`

Un principale con questa autorizzazione può creare una tabella di metadati o un collegamento a una risorsa nel Catalogo dati all'interno del database specificato.

Example

L'esempio seguente concede all'utente l'`datalake_user1`autorizzazione a creare tabelle nel `retail` database nell' AWS account 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
```

```
--permissions "CREATE_TABLE" --resource '{ "Database": {"Name":"retail"}}'
```

Quando un principale crea una tabella nel Data Catalog, tutti i permessi di Lake Formation sulla tabella vengono concessi al principale, con la possibilità di concedere tali autorizzazioni ad altri.

Sovvenzioni tra account

Se un account del proprietario del database concede `CREATE_TABLE` a un account destinatario e un utente dell'account destinatario crea correttamente una tabella nel database dell'account proprietario, si applicano le seguenti regole:

- Gli amministratori dell'utente e del data lake nell'account del destinatario dispongono di tutte le autorizzazioni di Lake Formation sulla tabella. Possono concedere le autorizzazioni sulla tabella ad altri responsabili del loro account. Non possono concedere autorizzazioni ai responsabili dell'account proprietario o di qualsiasi altro account.
- Gli amministratori di Data Lake dell'account proprietario possono concedere le autorizzazioni relative alla tabella ad altri responsabili del proprio account.

Autorizzazioni per la localizzazione dei dati

Quando tenti di creare una tabella che punti a una posizione Amazon S3, a seconda che tu disponga delle autorizzazioni per la localizzazione dei dati, l'`CREATE_TABLE`autorizzazione potrebbe non essere sufficiente per creare una tabella. È importante tenere a mente i seguenti tre casi.

Crea un caso d'uso della tabella	Autorizzazioni necessarie
La località specificata non è gestita da Lake Formation (non è registrata).	`CREATE_TABLE` è sufficiente.
La posizione specificata è gestita da Lake Formation (è registrata) e il database che lo contiene non ha alcuna proprietà di location o ha una proprietà di location che non sia un prefisso Amazon S3 della posizione della tabella.	`CREATE_TABLE` è richiesta più le autorizza zioni di localizzazione dei dati nella posizione specificata.
La posizione specificata è gestita da Lake Formation (è registrata) e il database contenent	`CREATE_TABLE` è sufficiente.

Crea un caso d'uso della tabella	Autorizzazioni necessarie
e ha una proprietà location che punta a una posizione registrata ed è un prefisso Amazon S3 della posizione della tabella.	

DATA_LOCATION_ACCESS

Autorizzazione	Concesso su questa risorsa	Inoltre, il beneficiario ha bisogno
DATA_LOCATION_ACCESS	Posizione di Amazon S3.	(Autorizzazioni Amazon S3 sulla posizione, che devono essere specificate dal ruolo utilizzato per registrare la posizione.)

Questa è l'unica autorizzazione per la localizzazione dei dati. Un principale con questa autorizzazione può creare un database o una tabella di metadati che punti alla posizione Amazon S3 specificata. La sede deve essere registrata. Un preside che dispone delle autorizzazioni di localizzazione dei dati su una sede dispone anche delle autorizzazioni di localizzazione sulle sedi dei figli.

Example

L'esempio seguente concede le autorizzazioni di localizzazione dei dati `s3://products/retail` all'utente `datalake_user1` nell'account 1111-2222-3333. AWS

```
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`non è necessario per interrogare o aggiornare i dati sottostanti. Questa autorizzazione si applica solo alla creazione di risorse del catalogo dati.

Per ulteriori informazioni sulle autorizzazioni per la localizzazione dei dati, vedereUnderlying data access control.

DELETE

Autorizzazione	Concesso su questa risorsa	Inoltre, il beneficiario ha bisogno
DELETE	TABLE	(Non sono necessarie autorizzazioni IAM aggiuntive se la posizione è registrata.)

Un principale con questa autorizzazione può eliminare i dati sottostanti nella posizione Amazon S3 specificata nella tabella. Il responsabile può anche visualizzare la tabella sulla console Lake Formation e recuperare informazioni sulla tabella con l'AWS GlueAPI.

Example

L'esempio seguente concede l'DELETEautorizzazione all'utente `datalake_user1` sulla tabella del database `inventory` `retail` nell' AWS account 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DELETE" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

Questa autorizzazione si applica solo ai dati in Amazon S3 e non ai dati in altri archivi dati come Amazon Relational Database Service (Amazon RDS).

DESCRIBE

Autorizzazione	Concesso su questa risorsa	Inoltre, il beneficiario ha bisogno
DESCRIBE	Link alle risorse della tabella	`glue:GetTable`
	Collegamento alle risorse del database	`glue:GetDatabase`
DESCRIBE	DATABASE	`glue:GetDatabase`
DESCRIBE	TABLE	`glue:GetTable`

Autorizzazione	Concesso su questa risorsa	Inoltre, il beneficiario ha bisogno

`DESCRIBE`	`LF-Tag`	`glue:GetTable`
		`glue:GetDatabase`
		`lakeformation:GetResourceLFTags`
		`lakeformation:ListLFTags`
		`lakeformation:GetLFTag`
		`lakeformation:SearchTablesByLFTags`
		`lakeformation:SearchDatabasesByLFTags`

Un principale con questa autorizzazione può visualizzare il database, la tabella o il link alla risorsa specificato. Non vengono concesse implicitamente altre autorizzazioni di Data Catalog e non vengono concesse implicitamente autorizzazioni di accesso ai dati. I database e le tabelle vengono visualizzati negli editor di query dei servizi integrati, ma non è possibile eseguire alcuna query su di essi a meno che non vengano concesse altre autorizzazioni di Lake Formation (ad esempio`SELECT`).

Ad esempio, un utente che ha un database può visualizzare il database e tutti i metadati del database (descrizione, posizione e così `DESCRIBE` via). Tuttavia, l'utente non può scoprire quali tabelle contiene il database e non può eliminare, modificare o creare tabelle nel database. Analogamente, un utente che `DESCRIBE` utilizza una tabella può visualizzare la tabella e i relativi metadati (descrizione, schema, posizione e così via), ma non può eliminare, modificare o eseguire query sulla tabella.

Di seguito sono riportate alcune regole aggiuntive per: `DESCRIBE`

- Se un utente dispone di altre autorizzazioni Lake Formation su un database, una tabella o un collegamento a una risorsa, `DESCRIBE` viene concessa implicitamente.

- Se un utente ha `SELECT` solo un sottoinsieme di colonne per una tabella (parziale`SELECT`), l'utente è limitato a visualizzare solo quelle colonne.
- Non puoi concedere `DESCRIBE` a un utente che dispone di una selezione parziale su una tabella. Al contrario, non è possibile specificare elenchi di inclusione o esclusione di colonne per le tabelle `DESCRIBE` consentite.

Example

L'esempio seguente concede all'utente `datalake_user1` l'`DESCRIBE`autorizzazione per il collegamento alle risorse della tabella nel database `inventory-link` `retail` nell' AWS account 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

Autorizzazione	Concesso su questa risorsa	Inoltre, il beneficiario ha bisogno
`DROP`	`DATABASE`	`glue:DeleteDatabase`
`DROP`	`TABLE`	`glue:DeleteTable`
`DROP`	`LF-Tag`	`lakeformation:DeleteLFTag`

`DROP`	Collegamento alla risorsa del database
	
	Collegamento alle risorse della tabella
		`glue:DeleteDatabase`
		
		`glue:DeleteTable`

Un principale con questa autorizzazione può eliminare un collegamento a un database, una tabella o una risorsa nel Catalogo dati. Non puoi concedere DROP su un database a un account o un'organizzazione esterni.

> **Warning**
>
> L'eliminazione di un database comporta l'eliminazione di tutte le tabelle del database.

Example

L'esempio seguente concede l'DROPautorizzazione all'utente del database `retail` nell' AWS account `datalake_user1` 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "DROP" --resource '{ "Database": {"Name":"retail"}}'
```

Example

L'esempio seguente concede DROP all'utente l'accesso alla tabella del database`datalake_user1`. `inventory` `retail`

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DROP" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

Example

L'esempio che segue concede DROP all'utente `datalake_user1` della tabella un link di risorse `inventory-link` nel database. `retail`

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "DROP" --resource '{ "Table": {"DatabaseName":"retail", "Name":"inventory-
link"}}'
```

INSERT

Autorizzazione	Concesso su questa risorsa	Inoltre, il beneficiario ha bisogno
INSERT	TABLE	(Non sono necessarie autorizzazioni IAM aggiuntive se la posizione è registrata.)

Un principale con questa autorizzazione può inserire, aggiornare e leggere i dati sottostanti nella posizione Amazon S3 specificata dalla tabella. Il responsabile può anche visualizzare la tabella nella console Lake Formation e recuperare informazioni sulla tabella con l'AWS GlueAPI.

Example

L'esempio seguente concede l'INSERTautorizzazione all'utente `datalake_user1` sulla tabella del database `inventory` `retail` nell' AWS account 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "INSERT" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

Questa autorizzazione si applica solo ai dati in Amazon S3 e non ai dati in altri archivi di dati come Amazon RDS.

SELECT

Autorizzazione	Concesso su questa risorsa	Inoltre, il beneficiario ha bisogno
SELECT	• TABLE	(Non sono necessarie autorizzazioni IAM aggiuntive se la posizione è registrata.)

Un principale con questa autorizzazione può visualizzare una tabella nel catalogo dati e interrogare i dati sottostanti in Amazon S3 nella posizione specificata dalla tabella. Il responsabile può visualizzare

la tabella nella console di Lake Formation e recuperare informazioni sulla tabella con l'AWS GlueAPI. Se il filtraggio delle colonne è stato applicato quando è stata concessa questa autorizzazione, il principale può visualizzare i metadati solo per le colonne incluse e può interrogare i dati solo dalle colonne incluse.

> **Note**
>
> È responsabilità del servizio di analisi integrato applicare il filtro delle colonne durante l'elaborazione di una query.

Example

L'esempio seguente concede l'`SELECT`autorizzazione all'utente `datalake_user1` sulla tabella del database `inventory` `retail` nell' AWS account 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "SELECT" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

Questa autorizzazione si applica solo ai dati in Amazon S3 e non ai dati in altri archivi di dati come Amazon RDS.

Puoi filtrare (limitare l'accesso a) colonne specifiche con un elenco di inclusione o un elenco di esclusione facoltativo. Un elenco di inclusione specifica le colonne a cui è possibile accedere. Un elenco di esclusione specifica le colonne a cui non è possibile accedere. In assenza di un elenco di inclusione o esclusione, tutte le colonne della tabella sono accessibili.

I risultati `glue:GetTable` restituiscono solo le colonne che il chiamante è autorizzato a visualizzare. I servizi integrati come Amazon Athena e Amazon Redshift rispettano gli elenchi di inclusione ed esclusione delle colonne.

Example

L'esempio seguente concede `SELECT` all'utente presente nella tabella l'`inventory`utilizzo di `datalake_user1` un elenco di inclusione.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
```

```
permissions "SELECT"  --resource '{ "TableWithColumns": {"DatabaseName":"retail",
 "Name":"inventory", "ColumnNames": ["prodcode","location","period","withdrawals"]}}'
```

Example

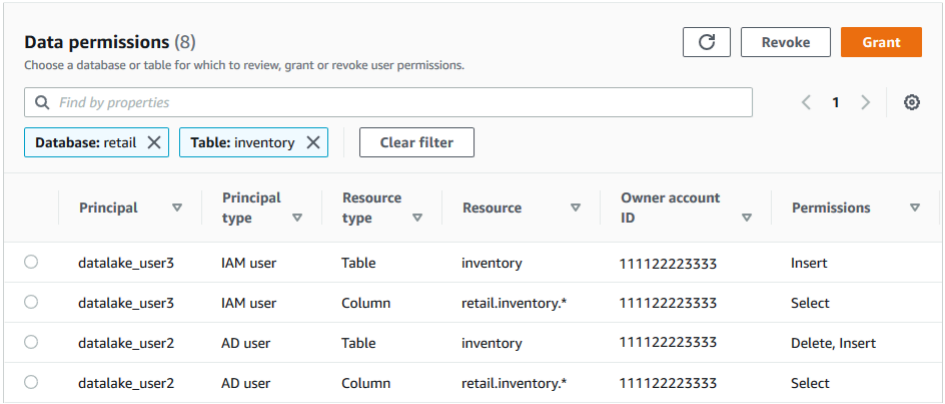
L'esempio successivo prevede l'utilizzo di SELECT un elenco di esclusione `inventory` nella tabella.

```
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"]}}}'
```

Le seguenti restrizioni si applicano all'SELECTautorizzazione:

- Quando si concedeSELECT, non è possibile includere l'opzione di concessione se viene applicato il filtro delle colonne.
- Non è possibile limitare il controllo dell'accesso alle colonne che sono chiavi di partizione.
- A un principale con l'SELECTautorizzazione su un sottoinsieme di colonne di una tabella non può essere concessa l'INSERTautorizzazioneALTER,DROP,DELETE, o per quella tabella. Analogamente, a un principale con l'INSERTautorizzazione ALTERDROP,DELETE, o su una tabella non può essere concessa l'SELECTautorizzazione con il filtro delle colonne.

L'SELECTautorizzazione appare sempre nella pagina Autorizzazioni dati della console Lake Formation come una riga separata. L'immagine seguente mostra che SELECT è concesso agli utenti `datalake_user2` e `datalake_user3` su tutte le colonne della `inventory` tabella.



Super

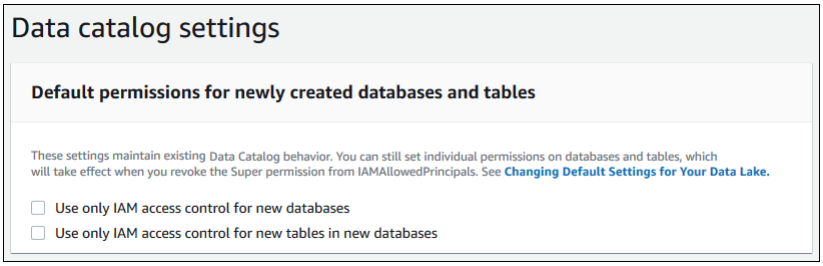
Autorizzazione	Concesso su questa risorsa	Anche il beneficiario ha bisogno
Super	DATABASE	glue:*Database*
Super	TABLE	glue:*Table*, glue:*Partition*

Questa autorizzazione consente a un principale di eseguire tutte le operazioni di Lake Formation supportate sul database o sulla tabella. Non puoi concedere Super un database a un account esterno.

Questa autorizzazione può coesistere con le altre autorizzazioni di Lake Formation. Ad esempio, puoi concedere le INSERT autorizzazioni SuperSELECT, e su una tabella di metadati. Il principale può quindi eseguire tutte le operazioni supportate sulla tabella. Quando si revocaSuper, i INSERT permessi SELECT and rimangono e il principale può eseguire solo operazioni di selezione e inserimento.

Invece di concedere Super a un singolo committente, puoi concederlo al gruppo. IAMAllowedPrincipals Il IAMAllowedPrincipals gruppo viene creato automaticamente e include tutti gli utenti e i ruoli IAM a cui è consentito l'accesso alle risorse del Data Catalog dalle politiche IAM. Quando Super viene concesso a IAMAllowedPrincipals una risorsa Data Catalog, l'accesso alla risorsa è effettivamente controllato esclusivamente dalle politiche IAM.

Puoi avere l'Superautorizzazione da concedere automaticamente IAMAllowedPrincipals per nuove risorse del catalogo sfruttando le opzioni nella pagina Impostazioni della console Lake Formation.



- Per concedere Super a IAMAllowedPrincipals tutti i nuovi database, seleziona Usa solo il controllo di accesso IAM per i nuovi database.

- Per concedere `Super` a `IAMAllowedPrincipals` tutte le nuove tabelle nei nuovi database, seleziona Usa solo il controllo di accesso IAM per le nuove tabelle nei nuovi database.

> **Note**
>
> Questa opzione fa sì che la casella di controllo Usa solo il controllo di accesso IAM per le nuove tabelle in questo database nella finestra di dialogo Crea database sia selezionata per impostazione predefinita. Non fa altro che questo. È la casella di controllo nella finestra di dialogo Crea database che abilita la concessione di `Super` a`IAMAllowedPrincipals`.

Queste opzioni della pagina Impostazioni sono abilitate per impostazione predefinita. Per ulteriori informazioni, consulta gli argomenti seguenti:

- the section called “Modifica delle impostazioni predefinite per il data lake”
- the section called “Aggiornamento delle autorizzazioni per AWS Glue i dati al modello Lake Formation”

ASSOCIATE

Autorizzazione	Concesso su questa risorsa	Inoltre, il beneficiario ha bisogno

`ASSOCIATE`	`LF-Tag`	`glue:GetDatabase`
		`glue:GetTable`
		`lakeformation:AddLFTagsToResource"`
		`lakeformation:RemoveLFTagsFromResource"`
		`lakeformation:GetResourceLFTags`
		`lakeformation:ListLFTags`
		`lakeformation:GetLFTag`

Autorizzazione	Concesso su questa risorsa	Inoltre, il beneficiario ha bisogno

		`lakeformation:SearchTablesByLFTags`
		
		`lakeformation:SearchDatabasesByLFTags`

Un principale con questa autorizzazione su un tag LF può assegnare il tag LF a una risorsa del catalogo dati. `ASSOCIATEDESCRIBE`Concedere concessioni implicite.

Example

Questo esempio concede all'utente l'`ASSOCIATE`autorizzazione per `datalake_user1` il tag LF con la chiave. `module` Concede le autorizzazioni per visualizzare e assegnare tutti i valori per quella chiave, come indicato dall'asterisco (*).

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --permissions "ASSOCIATE" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

Integrazione di IAM Identity Center

Con AWS IAM Identity Center, puoi connetterti ai provider di identità (IdPs) e gestire centralmente l'accesso per utenti e gruppi attraverso i servizi di AWS analisi. Puoi integrare provider di identità come Okta, Ping e Microsoft Entra ID (precedentemente Azure Active Directory) con IAM Identity Center per consentire agli utenti della tua organizzazione di accedere ai dati utilizzando un'esperienza di accesso singolo. IAM Identity Center supporta anche la connessione di altri provider di identità di terze parti.

Per ulteriori informazioni, consulta la sezione Provider di identità supportati nella Guida AWS IAM Identity Center per l'utente.

Puoi configurarlo AWS Lake Formation come applicazione abilitata in IAM Identity Center e gli amministratori del data lake possono concedere autorizzazioni granulari a utenti e gruppi autorizzati sulle risorse. AWS Glue Data Catalog

Gli utenti della tua organizzazione possono accedere a qualsiasi applicazione abilitata per Identity Center utilizzando il provider di identità dell'organizzazione e interrogare i set di dati utilizzando le autorizzazioni Lake Formation. Con questa integrazione, puoi gestire l'accesso ai AWS servizi, senza creare più ruoli IAM.

> **Note**
>
> La propagazione affidabile delle identità consente agli utenti e ai gruppi di appartenenza esistenti degli utenti di accedere ai dati di tutti AWS i servizi di analisi. Con la propagazione affidabile delle identità, un utente può accedere a un'applicazione e l'applicazione può trasmettere l'identità dell'utente nelle richieste di accesso ai dati nei servizi. AWS Non è necessario eseguire configurazioni di provider di identità specifiche del servizio o configurazioni di ruoli IAM. Per ulteriori informazioni, consulta Trusted Identity Propagation tra le applicazioni nella Guida per l'utente. AWS IAM Identity Center

Per le limitazioni, consulta Limitazioni dell'integrazione di IAM Identity Center.

Argomenti

- Prerequisiti
- Connessione di Lake Formation con IAM Identity Center
- Aggiornamento di un'integrazione con IAM Identity Center
- Eliminazione di una connessione Lake Formation con IAM Identity Center
- Concessione di autorizzazioni a utenti e gruppi

Prerequisiti

Di seguito sono riportati i prerequisiti per l'integrazione di IAM Identity Center con Lake Formation.

1. Abilita IAM Identity Center: abilitare IAM Identity Center è un prerequisito per supportare l'autenticazione e la propagazione delle identità.

2. Scegli la tua fonte di identità: dopo aver abilitato IAM Identity Center, devi disporre di un provider di identità per gestire utenti e gruppi. È possibile utilizzare la directory Identity Center integrata come origine di identità o utilizzare un IdP esterno, ad esempio Microsoft Entra ID o Okta.

   Per ulteriori informazioni, consulta Manage your identity source e Connect to a un provider di identità esterno nella Guida per l' AWS IAM Identity Center utente.

3. Crea un ruolo IAM: il ruolo che crea la connessione IAM Identity Center richiede le autorizzazioni per creare e modificare la configurazione dell'applicazione in Lake Formation e IAM Identity Center come nella seguente politica in linea.

   È necessario aggiungere le autorizzazioni in base alle migliori pratiche IAM. Le autorizzazioni specifiche sono illustrate nelle procedure che seguono. Per ulteriori informazioni, consulta Guida introduttiva a IAM Identity Center.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:CreateLakeFormationIdentityCenterConfiguration",
                "sso:CreateApplication",
                "sso:PutApplicationAssignmentConfiguration",
                "sso:PutApplicationAuthenticationMethod",
                "sso:PutApplicationGrant",
                "sso:PutApplicationAccessScope"
            ],
            "Resource": [
                "*"
            ]
        }
    ]
}
```

Le seguenti politiche in linea contengono autorizzazioni specifiche necessarie per visualizzare, aggiornare ed eliminare le proprietà dell'integrazione di Lake Formation con IAM Identity Center.

- Utilizza la seguente policy in linea per consentire a un ruolo IAM di visualizzare un'integrazione di Lake Formation con IAM Identity Center.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:DescribeLakeFormationIdentityCenterConfiguration",
                "sso:DescribeApplication"
            ],
            "Resource": [
                "*"
            ]
        }
    ]
}
```

- Utilizza la seguente policy in linea per consentire a un ruolo IAM di aggiornare un'integrazione di Lake Formation con IAM Identity Center.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:UpdateLakeFormationIdentityCenterConfiguration",
                "lakeformation:DescribeLakeFormationIdentityCenterConfiguration",
                "sso:DescribeApplication",
                "sso:UpdateApplication"
            ],
            "Resource": [
                "*"
            ]
        }
    ]
}
```

- Utilizza la seguente policy in linea per consentire a un ruolo IAM di eliminare un'integrazione di Lake Formation con IAM Identity Center.

```
{
```

```
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:DeleteLakeFormationIdentityCenterConfiguration",
                "sso:DeleteApplication"
            ],
            "Resource": [
                "*"
            ]
        }
    ]
}
```

- Per le autorizzazioni IAM necessarie per concedere o revocare le autorizzazioni del data lake per utenti e gruppi IAM Identity Center, consulta. Autorizzazioni IAM necessarie per concedere o revocare le autorizzazioni di Lake Formation

Descrizione delle autorizzazioni

- `lakeformation:CreateLakeFormationIdentityCenterConfiguration`— Crea la configurazione iDC di Lake Formation.
- `lakeformation:DescribeLakeFormationIdentityCenterConfiguration`— Descrive una configurazione iDC esistente.
- `lakeformation:DeleteLakeFormationIdentityCenterConfiguration`— Offre la possibilità di eliminare una configurazione iDC di Lake Formation esistente.
- `lakeformation:UpdateLakeFormationIdentityCenterConfiguration`— Usato per modificare una configurazione esistente di Lake Formation.
- `sso:CreateApplication`: utilizzato per creare un'applicazione Centro identità IAM.
- `sso:DeleteApplication`: utilizzato per eliminare un'applicazione Centro identità IAM.
- `sso:UpdateApplication`: utilizzato per aggiornare un'applicazione Centro identità IAM.
- `sso:PutApplicationGrant`: utilizzato per modificare le informazioni sull'emittente di token attendibile.
- `sso:PutApplicationAuthenticationMethod`— Garantisce l'accesso all'autenticazione di Lake Formation.

- `sso:GetApplicationGrant`: utilizzato per elencare informazioni sull'emittente di token attendibile.
- `sso:DeleteApplicationGrant`: elimina le informazioni sull'emittente di token attendibile.
- `sso:PutApplicationAccessScope`— Aggiunge o aggiorna l'elenco di destinazioni autorizzate per un ambito di accesso di IAM Identity Center per un'applicazione.
- `sso:PutApplicationAssignmentConfiguration`— Utilizzato per configurare il modo in cui gli utenti accedono a un'applicazione.

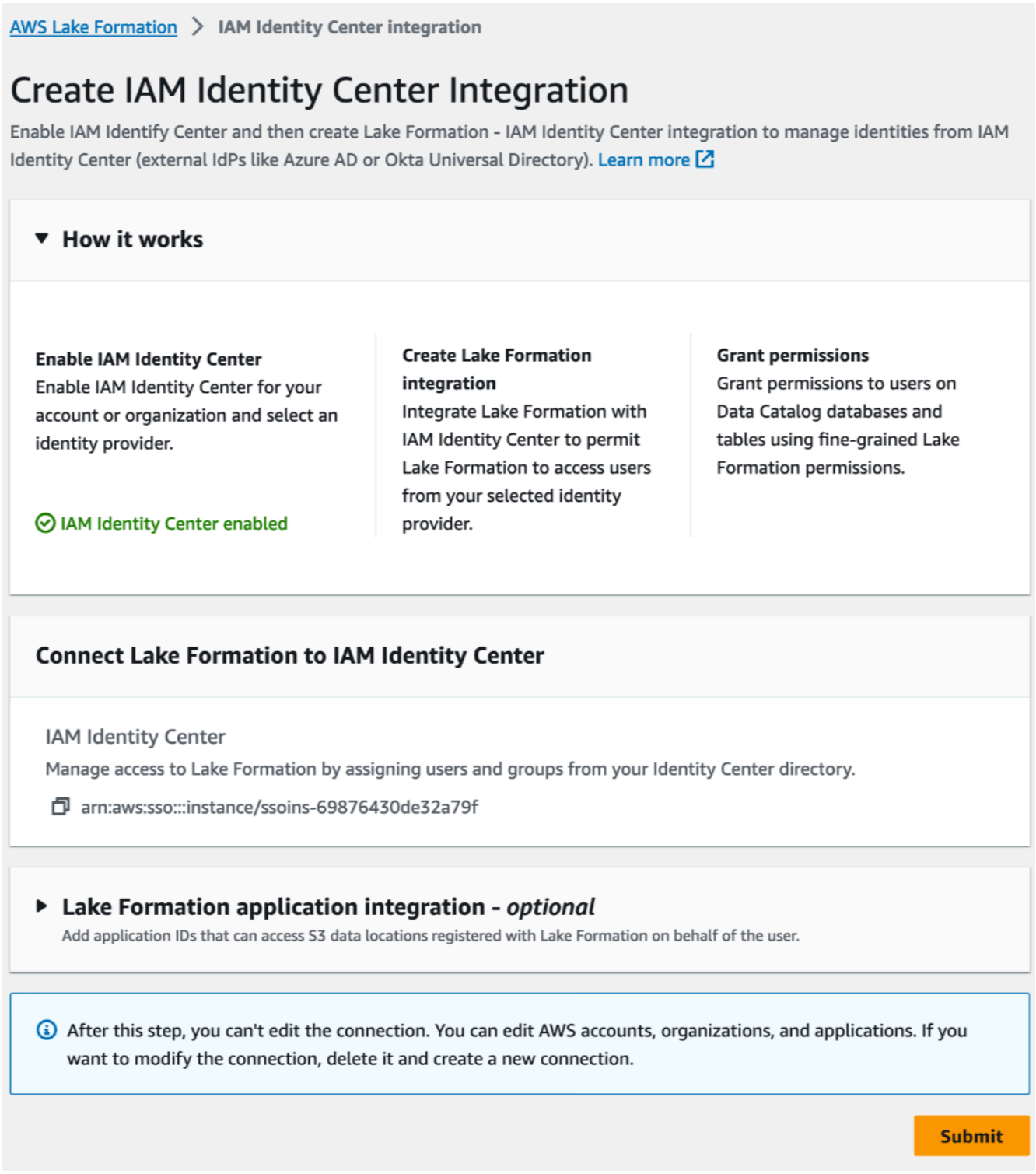
Connessione di Lake Formation con IAM Identity Center

Prima di poter utilizzare IAM Identity Center per gestire le identità e concedere l'accesso alle risorse del Data Catalog utilizzando Lake Formation, devi completare i seguenti passaggi. Puoi creare l'integrazione con IAM Identity Center utilizzando la console Lake Formation oppure AWS CLI.

AWS Management Console

Per connettere Lake Formation con IAM Identity Center

1. Accedi a e apri AWS Management Console la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.
2. Nel riquadro di navigazione a sinistra, seleziona IAM Identity Center integration.



3. (Facoltativo) Nella schermata di integrazione Create Lake Formation, specifica gli ARN delle applicazioni di terze parti che possono accedere ai dati nelle località Amazon S3 registrate con Lake Formation. Lake Formation fornisce credenziali temporanee limitate sotto forma di token alle sedi Amazon S3 registrate in base AWS STS alle autorizzazioni effettive, in modo che le applicazioni autorizzate possano accedere ai dati per conto degli utenti.

4. Selezionare Invia.

   Dopo che l'amministratore di Lake Formation ha completato i passaggi e creato l'integrazione, le proprietà di IAM Identity Center vengono visualizzate nella console di Lake Formation. Il completamento di queste attività rende Lake Formation un'applicazione abilitata per IAM Identity Center. Le proprietà della console includono lo stato dell'integrazione. Lo stato dell'integrazione indica `Success` quando è completata. Questo stato indica se la configurazione di IAM Identity Center è stata completata.

AWS CLI

- L'esempio seguente mostra come creare l'integrazione di Lake Formation con IAM Identity Center. Puoi anche specificare il `Status` (`ENABLED`,`DISABLED`) delle applicazioni.

```
aws lakeformation create-lake-formation-identity-center-configuration \
    --catalog-id <123456789012> \
    --instance-arn <arn:aws:sso:::instance/ssoins-112111f12ca1122p> \
    --external-filtering '{"AuthorizedTargets": ["<app arn1>", "<app arn2>"],
 "Status": "ENABLED"}'
```

- L'esempio seguente mostra come visualizzare un'integrazione di Lake Formation con IAM Identity Center.

```
aws lakeformation describe-lake-formation-identity-center-configuration
 --catalog-id <123456789012>
```

Aggiornamento di un'integrazione con IAM Identity Center

Dopo aver creato la connessione, puoi aggiungere applicazioni di terze parti per l'integrazione di IAM Identity Center da integrare con Lake Formation e ottenere l'accesso ai dati di Amazon S3 per conto degli utenti. Puoi anche rimuovere le applicazioni esistenti dall'integrazione con IAM Identity Center. È possibile aggiungere o rimuovere applicazioni utilizzando la console Lake Formation e utilizzando UpdateLakeFormationIdentityCenterConfigurationoperation. AWS CLI

> **Note**
>
> Dopo aver creato l'integrazione con IAM Identity Center, non puoi aggiornare l'istanzaARN.

AWS Management Console

Per aggiornare una connessione IAM Identity Center esistente con Lake Formation

1. Accedi a e apri AWS Management Console la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.
2. Nel riquadro di navigazione a sinistra, seleziona IAM Identity Center integration.
3. Seleziona Aggiungi nella pagina di integrazione di IAM Identity Center.
4. Nella schermata Aggiungi applicazioni, inserisci gli ID delle applicazioni di terze parti che desideri integrare con Lake Formation.
5. Selezionare Aggiungi.

AWS CLI

Puoi aggiungere o rimuovere applicazioni di terze parti per l'integrazione con IAM Identity Center eseguendo il AWS CLI comando seguente. Quando imposti lo stato del filtro esterno su`ENABLED`, consente a IAM Identity Center di fornire la gestione delle identità per le applicazioni di terze parti per accedere ai dati gestiti da Lake Formation. Puoi anche abilitare o disabilitare l'integrazione di IAM Identity Center impostando lo stato dell'applicazione.

```
aws lakeformation update-lake-formation-identity-center-configuration \
 --external-filtering '{"AuthorizedTargets": ["<app arn1>", "<app arn2>"], "Status":
 "ENABLED"}'\
 --application-status ENABLED
```

Eliminazione di una connessione Lake Formation con IAM Identity Center

Se desideri eliminare un'integrazione esistente di IAM Identity Center, puoi farlo utilizzando la console o l'DeleteLakeFormationIdentityCenterConfigurationoperazione Lake Formation. AWS CLI

AWS Management Console

Per eliminare una connessione IAM Identity Center esistente con Lake Formation

1. Accedi a e apri AWS Management Console la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.
2. Nel riquadro di navigazione a sinistra, seleziona IAM Identity Center integration.
3. Seleziona Elimina nella pagina di integrazione di IAM Identity Center.
4. Nella schermata di conferma dell'integrazione, conferma l'azione e seleziona Elimina.

AWS CLI

Puoi eliminare l'integrazione con IAM Identity Center eseguendo il AWS CLI comando seguente.

```
aws lakeformation delete-lake-formation-identity-center-configuration \
    --catalog-id <123456789012>
```

Concessione di autorizzazioni a utenti e gruppi

L'amministratore del data lake può concedere autorizzazioni a utenti e gruppi di IAM Identity Center sulle risorse del Data Catalog (database, tabelle e viste) per consentire un facile accesso ai dati. Per concedere o revocare le autorizzazioni del data lake, il concedente richiede le autorizzazioni per le seguenti azioni di IAM Identity Center.

- DescribeUser
- DescribeGroup
- DescribeInstance

Puoi concedere le autorizzazioni utilizzando la console Lake Formation, l'API o il AWS CLI.

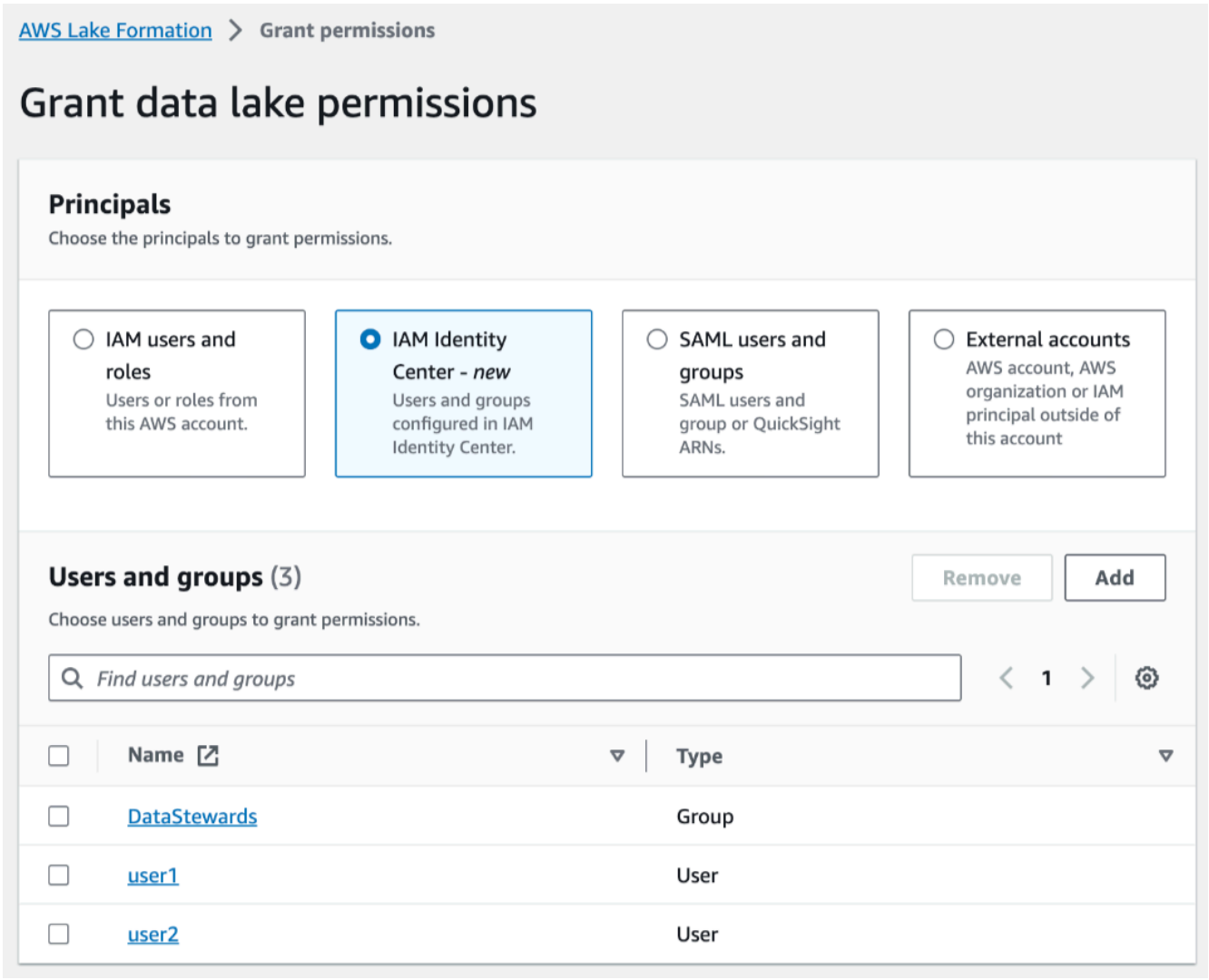
Per ulteriori informazioni sulla concessione delle autorizzazioni, consulta. the section called “Concessione e revoca delle autorizzazioni di Data Catalog”

> **Note**
>
> Puoi concedere autorizzazioni solo sulle risorse del tuo account. Per assegnare a catena le autorizzazioni a utenti e gruppi sulle risorse condivise con te, devi utilizzare AWS RAM le condivisioni di risorse.

AWS Management Console

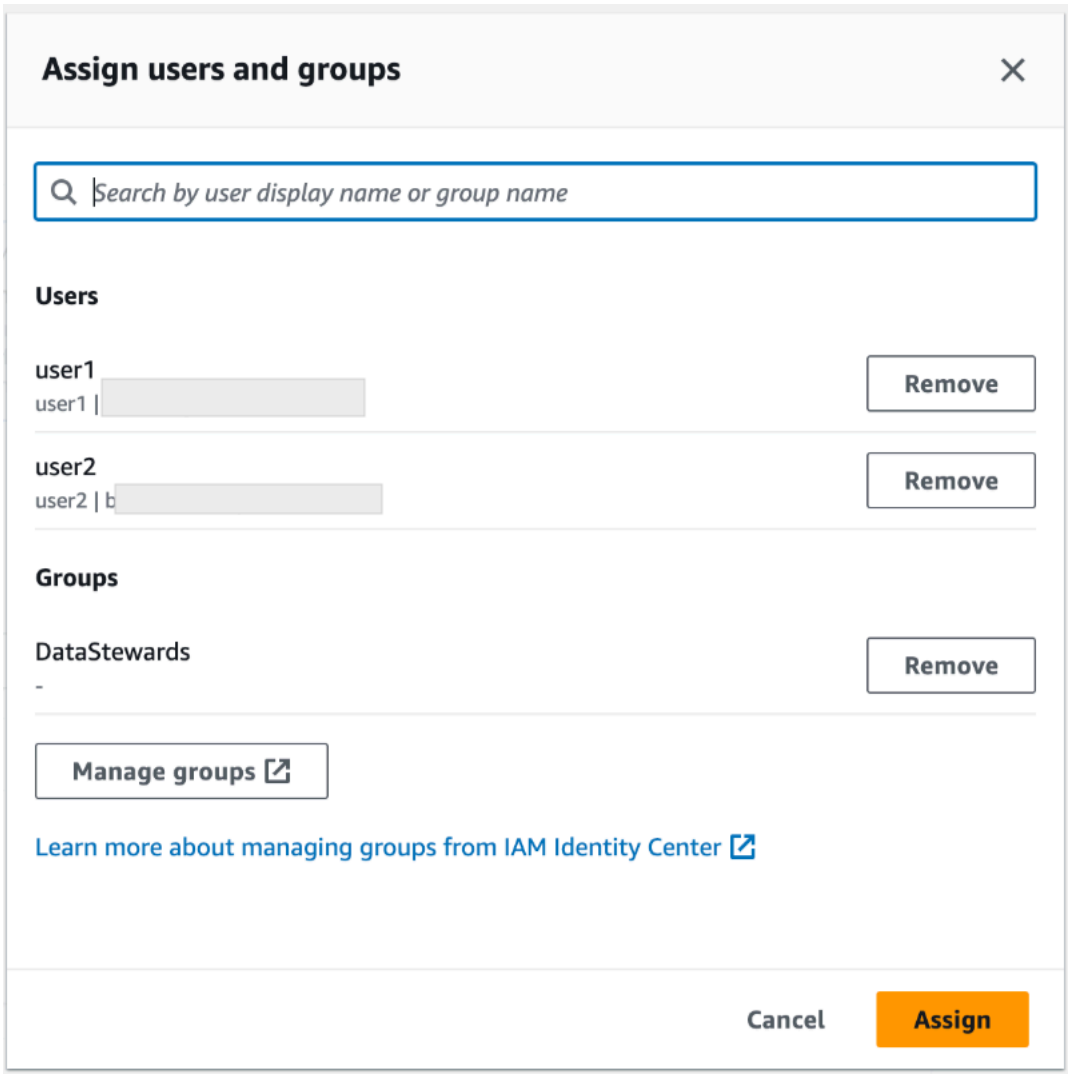
Per concedere autorizzazioni a utenti e gruppi

1. Accedi a e apri AWS Management Console la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.
2. Seleziona le autorizzazioni del data lake in Autorizzazioni nella console Lake Formation.
3. Seleziona Concedi.
4. Nella pagina Concedi le autorizzazioni del data lake, scegli, utenti e gruppi SSM.
5. Seleziona Aggiungi per scegliere gli utenti e i gruppi a cui concedere le autorizzazioni.



6. Nella schermata Assegna utenti e gruppi, scegli gli utenti e/o i gruppi a cui concedere le autorizzazioni.

   Seleziona Assegna.



7. Quindi, scegli il metodo per concedere le autorizzazioni.

   Per istruzioni sulla concessione delle autorizzazioni utilizzando il metodo delle risorse denominate, vedere. Concessione delle autorizzazioni per il data lake utilizzando il metodo di risorsa denominato

   Per istruzioni su come concedere l'autorizzazione utilizzando i tag LF, vedere. Concessione delle autorizzazioni per il data lake utilizzando il metodo LF-TBAC

8. Scegliete le risorse del Data Catalog per le quali desiderate concedere le autorizzazioni.
9. Scegli le autorizzazioni del Data Catalog da concedere.
10. Seleziona Concedi.

AWS CLI

L'esempio seguente mostra come concedere l'`SELECT`autorizzazione utente di IAM Identity Center su una tabella.

```
aws lakeformation grant-permissions \
--principal DataLakePrincipalIdentifier=arn:aws:identitystore:::user/<UserId> \
--permissions "SELECT" \
--resource '{ "Table": { "DatabaseName": "retail", "TableWildcard": {} } }'
```

Per eseguire il recupero `UserId` da IAM Identity Center, consulta il GetUserIdfunzionamento nel riferimento all'API IAM Identity Center.

Aggiungere una posizione Amazon S3 al tuo data lake

Per aggiungere una posizione Amazon Simple Storage Service (Amazon S3) come spazio di archiviazione nel tuo data lake, devi registrare la posizione con. AWS Lake Formation Puoi quindi utilizzare le autorizzazioni di Lake Formation per un controllo granulare degli accessi agli AWS Glue Data Catalog oggetti che puntano a questa posizione e ai dati sottostanti nella posizione.

Lake Formation consente inoltre di registrare una posizione dei dati in modalità di accesso ibrida e offre la flessibilità necessaria per abilitare selettivamente le autorizzazioni di Lake Formation per database e tabelle nel Data Catalog. Con la modalità di accesso Hybrid, ora disponi di un percorso incrementale che ti consente di impostare le autorizzazioni di Lake Formation per un set specifico di utenti senza interrompere le politiche di autorizzazione di altri utenti o carichi di lavoro esistenti.

Per ulteriori informazioni sulla configurazione della modalità di accesso ibrida, vedere Modalità di accesso ibrida

Quando registri una posizione, vengono registrati quel percorso Amazon S3 e tutte le cartelle in quel percorso.

Ad esempio, supponiamo di avere un'organizzazione dei percorsi Amazon S3 come la seguente:

`/mybucket/accounting/sales/`

Se ti registri`S3://mybucket/accounting`, anche la `sales` cartella è registrata e sotto la gestione di Lake Formation.

Per ulteriori informazioni sulla registrazione delle sedi, vedereUnderlying data access control.

> **Note**
>
> Le autorizzazioni Lake Formation sono consigliate per i dati strutturati (disposti in tabelle con righe e colonne). Se i tuoi dati contengono dati non strutturati basati su oggetti, valuta la possibilità di utilizzare l'autorizzazione IAM per Amazon S3 per gestire l'accesso ai dati.

Argomenti

Requisiti per i ruoli utilizzati per registrare le sedi

È necessario specificare un ruolo AWS Identity and Access Management (IAM) quando si registra una sede Amazon Simple Storage Service (Amazon S3). AWS Lake Formation assume quel ruolo quando accede ai dati in quella posizione.

È possibile utilizzare uno dei seguenti tipi di ruolo per registrare una posizione:

- Il ruolo legato ai servizi di Lake Formation. Questo ruolo concede le autorizzazioni necessarie sulla sede. L'utilizzo di questo ruolo è il modo più semplice per registrare la posizione. Per ulteriori informazioni, consulta Utilizzo di ruoli collegati ai servizi per Lake Formation.
- Un ruolo definito dall'utente. Utilizza un ruolo definito dall'utente quando devi concedere più autorizzazioni rispetto a quelle fornite dal ruolo collegato al servizio.

  È necessario utilizzare un ruolo definito dall'utente nelle seguenti circostanze:

  - Quando si registra una sede in un altro account.

    Per ulteriori informazioni, consultare the section called “Registrazione di una sede Amazon S3 in un altro account AWS” e the section called “Registrazione di una posizione Amazon S3 crittografata tra più account AWS”.

- Se hai utilizzato una CMK AWS gestita (`aws/s3`) per crittografare la posizione Amazon S3.

  Per ulteriori informazioni, consulta Registrazione di una posizione Amazon S3 crittografata.
- Se prevedi di accedere alla posizione utilizzando Amazon EMR.

  Se hai già registrato una sede con il ruolo collegato al servizio e desideri iniziare ad accedervi con Amazon EMR, devi annullare la registrazione della sede e registrarla nuovamente con un ruolo definito dall'utente. Per ulteriori informazioni, consulta the section called “Annullamento della registrazione di una sede Amazon S3”.

Utilizzo di ruoli collegati ai servizi per Lake Formation

AWS Lake Formationutilizza un ruolo collegato al AWS Identity and Access Management servizio (IAM). Un ruolo collegato ai servizi è un tipo unico di ruolo IAM collegato direttamente a Lake Formation. Il ruolo collegato al servizio è predefinito da Lake Formation e include tutte le autorizzazioni richieste dal servizio per chiamare altri AWS servizi per tuo conto.

Un ruolo collegato al servizio semplifica la configurazione di Lake Formation perché non è necessario creare un ruolo e aggiungere manualmente le autorizzazioni necessarie. Lake Formation definisce le autorizzazioni del suo ruolo collegato ai servizi e, se non diversamente definito, solo Lake Formation può assumerne i ruoli. Le autorizzazioni definite includono la policy di attendibilità e la policy delle autorizzazioni che non può essere collegata a nessun'altra entità IAM.

Questo ruolo collegato al servizio prevede che i seguenti servizi assumano il ruolo:

- `lakeformation.amazonaws.com`

Autorizzazioni di ruolo collegate al servizio per Lake Formation

Lake Formation utilizza il ruolo collegato al servizio denominato. `AWSServiceRoleForLakeFormationDataAccess` Questo ruolo fornisce una serie di autorizzazioni Amazon Simple Storage Service (Amazon S3) che consentono al servizio integrato Lake Formation (ad esempio) di accedere alle sedi Amazon Athena registrate. Quando registri una posizione di data lake, devi fornire un ruolo con le autorizzazioni di lettura/scrittura di Amazon S3 richieste in quella posizione. Invece di creare un ruolo con le autorizzazioni richieste di Amazon S3, puoi utilizzare questo ruolo collegato al servizio.

La prima volta che nomini il ruolo collegato al servizio come ruolo con cui registrare un percorso, il ruolo collegato al servizio e una nuova policy IAM vengono creati per tuo conto. Lake Formation

aggiunge il percorso alla politica in linea e la associa al ruolo collegato ai servizi. Quando registri percorsi successivi con il ruolo collegato al servizio, Lake Formation aggiunge il percorso alla policy esistente.

Dopo aver effettuato l'accesso come amministratore del data lake, registra una posizione di data lake. Quindi, nella console IAM, cerca il ruolo `AWSServiceRoleForLakeFormationDataAccess` e visualizza le policy allegate.

Ad esempio, dopo aver registrato la località`s3://my-kinesis-test/logs`, Lake Formation crea la seguente politica in linea e la allega a. `AWSServiceRoleForLakeFormationDataAccess`

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "LakeFormationDataAccessPermissionsForS3",
            "Effect": "Allow",
            "Action": [
                "s3:PutObject",
                "s3:GetObject",
                "s3:DeleteObject",
                "s3:AbortMultipartUpload",
                "s3:ListMultipartUploadParts"
            ],
            "Resource": [
                "arn:aws:s3:::my-kinesis-test/logs/*"
            ]
        },
        {
            "Sid": "LakeFormationDataAccessPermissionsForS3ListBucket",
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket",
                "s3:ListBucketMultipartUploads"
            ],
            "Resource": [
                "arn:aws:s3:::my-kinesis-test"
            ]
        }
    ]
}
```

Le seguenti autorizzazioni sono necessarie per poter registrare le sedi con questo ruolo collegato al servizio:

- `iam:CreateServiceLinkedRole`
- `iam:PutRolePolicy`

L'amministratore del data lake dispone in genere di queste autorizzazioni.

Di seguito sono riportati i requisiti per un ruolo definito dall'utente:

- Quando crei il nuovo ruolo, nella pagina Crea ruolo della console IAM, scegli AWS service, quindi in Choose a use case scegli Lake Formation.

  Se crei il ruolo utilizzando un percorso diverso, assicurati che il ruolo abbia una relazione di fiducia con`lakeformation.amazonaws.com`. Per ulteriori informazioni, vedere Modifica di una politica di attendibilità dei ruoli (console).
- Il ruolo deve avere relazioni di fiducia con le seguenti entità:
  - `glue.amazonaws.com`
  - `lakeformation.amazonaws.com`

  Per ulteriori informazioni, vedere Modifica di una politica di attendibilità dei ruoli (console).
- Il ruolo deve avere una politica in linea che conceda le autorizzazioni di lettura/scrittura di Amazon S3 sulla posizione. Di seguito è riportata una politica tipica.

```
{
    "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"
            ]
        }
    ]
}
```

- L'amministratore del data lake che registra la posizione deve disporre dell'`iam:PassRole`autorizzazione per il ruolo.

  Di seguito è riportata una politica in linea che concede questa autorizzazione. <account-id>Sostituiscilo con un numero di AWS account valido e <role-name>sostituiscilo con il nome del ruolo.

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

- Per consentire a Lake Formation di aggiungere log in CloudWatch Logs e pubblicare metriche, aggiungi la seguente politica in linea.

> **Note**
>
> La scrittura su CloudWatch Logs comporta un costo.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "Sid1",
            "Effect": "Allow",
            "Action": [
                "logs:CreateLogStream",
                "logs:CreateLogGroup",
                "logs:PutLogEvents"
            ],
            "Resource": [
                 "arn:aws:logs:<region>:<account-id>:log-group:/aws-lakeformation-
acceleration/*",
                 "arn:aws:logs:<region>:<account-id>:log-group:/aws-lakeformation-
acceleration/*:log-stream:*"
            ]
        }
    ]
}
```

Registrazione di una sede Amazon S3

È necessario specificare un ruolo AWS Identity and Access Management (IAM) quando si registra una sede Amazon Simple Storage Service (Amazon S3). Lake Formation assume questo ruolo quando concede credenziali temporanee ai AWS servizi integrati che accedono ai dati in quella posizione.

> **Important**
>
> Evita di registrare un bucket Amazon S3 con Requester pay abilitato. Per i bucket registrati con Lake Formation, il ruolo utilizzato per registrare il bucket viene sempre visualizzato come richiedente. Se al bucket si accede da un altro AWS account, al proprietario del bucket viene addebitato l'accesso ai dati se il ruolo appartiene allo stesso account del proprietario del bucket.

Puoi utilizzare la AWS Lake Formation console, l'API Lake Formation o AWS Command Line Interface (AWS CLI) per registrare una sede Amazon S3.

Prima di iniziare

Rivedi i requisiti per il ruolo utilizzato per registrare la sede.

Per registrare una posizione (console)

> **Important**
>
> Le seguenti procedure presuppongono che la posizione Amazon S3 si trovi nello stesso AWS account del Data Catalog e che i dati in essa contenuti non siano crittografati. Altre sezioni di questo capitolo riguardano la registrazione tra account e la registrazione di posizioni crittografate.

1. Apri la AWS Lake Formation console all'indirizzo https://console.aws.amazon.com/lakeformation/. Accedi come amministratore del data lake o come utente con l'autorizzazione `lakeformation:RegisterResource` IAM.
2. Nel riquadro di navigazione, in Register and Ingest, seleziona Data lake locations.
3. Scegli Registra posizione, quindi scegli Sfoglia per selezionare un percorso Amazon Simple Storage Service (Amazon S3).
4. (Facoltativo, ma fortemente consigliato) Seleziona Rivedi le autorizzazioni della sede per visualizzare un elenco di tutte le risorse esistenti nella posizione Amazon S3 selezionata e le relative autorizzazioni.

   La registrazione della località selezionata potrebbe consentire agli utenti di Lake Formation di accedere ai dati già presenti in quella posizione. La visualizzazione di questo elenco ti aiuta a garantire che i dati esistenti rimangano sicuri.
5. Per il ruolo IAM, scegli il ruolo `AWSServiceRoleForLakeFormationDataAccess` collegato al servizio (predefinito) o un ruolo IAM personalizzato che soddisfi i requisiti di. the section called “Requisiti per i ruoli utilizzati per registrare le sedi”

   Puoi aggiornare una posizione registrata o altri dettagli solo quando la registri utilizzando un ruolo IAM personalizzato. Per modificare una sede registrata utilizzando un ruolo collegato al servizio, è necessario annullare la registrazione della posizione e registrarla nuovamente.
6. Scegli l'opzione Enable Data Catalog Federation per consentire a Lake Formation di assumere un ruolo e vendere credenziali temporanee ai AWS servizi integrati per accedere alle tabelle nei database federati. Se una posizione è registrata con Lake Formation e desideri utilizzare la

stessa posizione per una tabella in un database federato, devi registrare la stessa posizione con l'opzione Enable Data Catalog Federation.

7. Scegli la modalità di accesso ibrida per non abilitare le autorizzazioni di Lake Formation per impostazione predefinita. Quando registri una posizione Amazon S3 in modalità di accesso ibrido, puoi abilitare le autorizzazioni Lake Formation optando per i principali database e tabelle in quella posizione.

   Per ulteriori informazioni sulla configurazione della modalità di accesso ibrida, consulta. Modalità di accesso ibrida

8. Seleziona Registra posizione.

Per registrare una sede (AWS CLI)

1. Registra una nuova sede con Lake Formation

   Questo esempio utilizza un ruolo collegato al servizio per registrare la posizione. È possibile utilizzare l'`--role-arn`argomento invece per fornire il proprio ruolo.

   <s3-path>Sostituiscilo con un percorso Amazon S3 valido, un numero di account con un AWS account valido e <s3-access-role>con un ruolo IAM che dispone delle autorizzazioni per registrare una posizione dati.

   > **Note**
   >
   > Non puoi modificare le proprietà di una posizione registrata se è registrata utilizzando un ruolo collegato al servizio.

```
aws lakeformation register-resource \
 --resource-arn arn:aws:s3:::<s3-path> \
 --use-service-linked-role
```

   L'esempio seguente utilizza un ruolo personalizzato per registrare la posizione.

```
aws lakeformation register-resource \
 --resource-arn arn:aws:s3:::<s3-path> \
 --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role>
```

2. Per aggiornare una località registrata con Lake Formation

   Puoi modificare una sede registrata solo se è registrata utilizzando un ruolo IAM personalizzato. Per una sede registrata con un ruolo collegato al servizio, è necessario annullare la registrazione della posizione e registrarla nuovamente. Per ulteriori informazioni, consulta the section called "Annullamento della registrazione di una sede 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. Registra una posizione dati in modalità di accesso ibrida con federazione

   ```
   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
   ```

Per ulteriori informazioni, consulta Funzionamento RegisterResourcedelle API.

> **Note**
>
> Dopo aver registrato una posizione Amazon S3, qualsiasi AWS Glue tabella che punta alla posizione (o a una delle sue sedi secondarie) restituirà il valore del `IsRegisteredWithLakeFormation` parametro come `true` nella chiamata. `GetTable` È noto che le operazioni dell'API Data Catalog, come `GetTables` e `SearchTables` non, aggiornano il valore del `IsRegisteredWithLakeFormation` parametro e restituiscono il valore predefinito, che è falso. Si consiglia di utilizzare l'`GetTable`API per visualizzare il valore corretto del `IsRegisteredWithLakeFormation` parametro.

Registrazione di una posizione Amazon S3 crittografata

Lake Formation si integra con AWS Key Management Service(AWS KMS) per consentirti di configurare più facilmente altri servizi integrati per crittografare e decrittografare i dati nelle sedi Amazon Simple Storage Service (Amazon S3).

Entrambi sono gestiti dal cliente e sono supportati. AWS KMS keys Chiavi gestite da AWS Attualmente, la crittografia/decrittografia lato client è supportata solo con Athena.

È necessario specificare un ruolo AWS Identity and Access Management (IAM) quando si registra una sede Amazon S3. Per le sedi Amazon S3 crittografate, il ruolo deve disporre dell'autorizzazione per crittografare e decrittografare i dati con o la politica della AWS KMS key chiave KMS deve concedere le autorizzazioni sulla chiave del ruolo.

> **Important**
>
> Evita di registrare un bucket Amazon S3 con Requester pay abilitato. Per i bucket registrati con Lake Formation, il ruolo utilizzato per registrare il bucket viene sempre visualizzato come richiedente. Se al bucket si accede da un altro AWS account, al proprietario del bucket viene addebitato l'accesso ai dati se il ruolo appartiene allo stesso account del proprietario del bucket.

Il modo più semplice per registrare la sede è utilizzare il ruolo collegato al servizio Lake Formation. Questo ruolo concede le autorizzazioni di lettura/scrittura richieste sulla posizione. È inoltre possibile utilizzare un ruolo personalizzato per registrare la posizione, a condizione che soddisfi i requisiti di. the section called “Requisiti per i ruoli utilizzati per registrare le sedi”

> **Important**
>
> Se hai usato un Chiave gestita da AWS (`aws/s3`) per crittografare la posizione Amazon S3, non puoi utilizzare il ruolo collegato al servizio Lake Formation. È necessario utilizzare un ruolo personalizzato e aggiungere le autorizzazioni IAM sulla chiave del ruolo. I dettagli sono forniti più avanti in questa sezione.

Le seguenti procedure spiegano come registrare una posizione Amazon S3 crittografata con una chiave gestita dal cliente o un. Chiave gestita da AWS

- Registrazione di una posizione crittografata con una chiave gestita dal cliente
- Registrazione di una posizione crittografata con un Chiave gestita da AWS

Prima di iniziare

Esamina i requisiti per il ruolo utilizzato per registrare la sede.

Per registrare una posizione Amazon S3 crittografata con una chiave gestita dal cliente

> **Note**
>
> Se la chiave KMS o la posizione Amazon S3 non si trovano AWS nello stesso account del Data Catalog, segui invece le istruzioni riportate the section called “Registrazione di una posizione Amazon S3 crittografata tra più account AWS” in.

1. Apri la AWS KMS console all'indirizzo https://console.aws.amazon.com/kms e accedi come utente amministrativo AWS Identity and Access Management (IAM) o come utente che può modificare la politica chiave della chiave KMS utilizzata per crittografare la posizione.
2. Nel riquadro di navigazione, scegli Customer managed keys, quindi scegli il nome della chiave KMS desiderata.
3. Nella pagina dei dettagli della chiave KMS, scegli la scheda Politica chiave, quindi esegui una delle seguenti operazioni per aggiungere il tuo ruolo personalizzato o il ruolo collegato al servizio Lake Formation come utente chiave KMS:

   - Se viene visualizzata la visualizzazione predefinita (con le sezioni Amministratori chiave, Eliminazione chiavi, Utenti chiave e Altri AWS account), nella sezione Utenti chiave,

aggiungi il tuo ruolo personalizzato o il ruolo collegato al servizio Lake Formation. `AWSServiceRoleForLakeFormationDataAccess`

- Se viene visualizzata la policy chiave (JSON), modifica la policy per aggiungere il ruolo personalizzato o il ruolo collegato `AWSServiceRoleForLakeFormationDataAccess` al servizio Lake Formation all'oggetto «Consenti l'uso della chiave», come mostrato nell'esempio seguente.

> **Note**
>
> Se quell'oggetto manca, aggiungilo con le autorizzazioni mostrate nell'esempio. L'esempio utilizza il ruolo collegato al servizio.

```
        ...
        {
            "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. Apri la AWS Lake Formation console all'indirizzo https://console.aws.amazon.com/lakeformation/. Accedi come amministratore del data lake o come utente con l'autorizzazione `lakeformation:RegisterResource` IAM.
5. Nel riquadro di navigazione, in Register and Ingest, scegli Data lake locations.

6. Scegli Registra posizione, quindi scegli Sfoglia per selezionare un percorso Amazon Simple Storage Service (Amazon S3).
7. (Facoltativo, ma fortemente consigliato) Scegli Rivedi le autorizzazioni della sede per visualizzare un elenco di tutte le risorse esistenti nella posizione Amazon S3 selezionata e le relative autorizzazioni.

   La registrazione della località selezionata potrebbe consentire agli utenti di Lake Formation di accedere ai dati già presenti in quella posizione. La visualizzazione di questo elenco ti aiuta a garantire che i dati esistenti rimangano sicuri.
8. Per il ruolo IAM, scegli il ruolo `AWSServiceRoleForLakeFormationDataAccess` collegato al servizio (predefinito) o il ruolo personalizzato che soddisfa i. the section called “Requisiti per i ruoli utilizzati per registrare le sedi”
9. Scegli Registra posizione.

Per ulteriori informazioni sul ruolo collegato al servizio, consulta Autorizzazioni di ruolo collegate al servizio per Lake Formation.

Per registrare una posizione Amazon S3 crittografata con un Chiave gestita da AWS

> **Important**
>
> Se la sede Amazon S3 non si trova nello stesso AWS account del Data Catalog, segui invece le istruzioni riportate inthe section called “Registrazione di una posizione Amazon S3 crittografata tra più account AWS”.

1. Crea un ruolo IAM da utilizzare per registrare la posizione. Assicurati che soddisfi i requisiti elencati inthe section called “Requisiti per i ruoli utilizzati per registrare le sedi”.
2. Aggiungi la seguente politica in linea al ruolo. Concede le autorizzazioni sulla chiave del ruolo. La `Resource` specifica deve indicare l'Amazon Resource Name (ARN) di. Chiave gestita da AWSÈ possibile ottenere l'ARN dalla AWS KMS console. Per ottenere l'ARN corretto, assicurati di accedere alla AWS KMS console con lo stesso AWS account e la stessa regione utilizzati per crittografare la Chiave gestita da AWS posizione.

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

```
    {
      "Effect": "Allow",
      "Action": [
        "kms:Encrypt",
        "kms:Decrypt",
        "kms:ReEncrypt*",
        "kms:GenerateDataKey*",
        "kms:DescribeKey"
      ],
      "Resource": "<Chiave gestita da AWS ARN>"
    }
  ]
}
```

3. Apri la AWS Lake Formation console all'indirizzo https://console.aws.amazon.com/lakeformation/. Accedi come amministratore del data lake o come utente con l'autorizzazione `lakeformation:RegisterResource` IAM.
4. Nel riquadro di navigazione, in Register and Ingest, scegli Data lake locations.
5. Scegli Registra posizione, quindi scegli Sfoglia per selezionare un percorso Amazon S3.
6. (Facoltativo, ma fortemente consigliato) Scegli Rivedi le autorizzazioni della sede per visualizzare un elenco di tutte le risorse esistenti nella posizione Amazon S3 selezionata e le relative autorizzazioni.

   La registrazione della località selezionata potrebbe consentire agli utenti di Lake Formation di accedere ai dati già presenti in quella posizione. La visualizzazione di questo elenco ti aiuta a garantire che i dati esistenti rimangano sicuri.
7. Per il ruolo IAM, scegli il ruolo che hai creato nella fase 1.
8. Scegli Registra posizione.

Registrazione di una sede Amazon S3 in un altro account AWS

AWS Lake Formation consente di registrare sedi Amazon Simple Storage Service (Amazon S3) su più account. AWS Ad esempio, se si AWS Glue Data Catalog trova nell'account A, un utente nell'account A può registrare un bucket Amazon S3 nell'account B.

La registrazione di un bucket Amazon S3 AWS nell'account B utilizzando AWS Identity and Access Management un ruolo (IAM) AWS nell'account A richiede le seguenti autorizzazioni:

- Il ruolo nell'account A deve concedere le autorizzazioni sul bucket dell'account B.

- La politica del bucket nell'account B deve concedere le autorizzazioni di accesso al ruolo nell'account A.

> **Important**
>
> Evita di registrare un bucket Amazon S3 con Requester pay abilitato. Per i bucket registrati con Lake Formation, il ruolo utilizzato per registrare il bucket viene sempre visualizzato come richiedente. Se al bucket si accede da un altro AWS account, al proprietario del bucket viene addebitato l'accesso ai dati se il ruolo appartiene allo stesso account del proprietario del bucket.
>
> Non puoi utilizzare il ruolo collegato al servizio Lake Formation per registrare una sede in un altro account. È invece necessario utilizzare un ruolo definito dall'utente. Il ruolo deve soddisfare i requisiti dithe section called “Requisiti per i ruoli utilizzati per registrare le sedi”. Per ulteriori informazioni sul ruolo collegato al servizio, consulta Autorizzazioni di ruolo collegate al servizio per Lake Formation.

Prima di iniziare

Esamina i requisiti per il ruolo utilizzato per registrare la sede.

Per registrare una sede in un altro AWS account

> **Note**
>
> Se la posizione è crittografata, segui the section called “Registrazione di una posizione Amazon S3 crittografata tra più account AWS” invece le istruzioni riportate in.

La procedura seguente presuppone che un titolare dell'account 1111-2222-3333, che contiene il Data Catalog, desideri registrare il `awsexamplebucket1` bucket Amazon S3, che si trova nell'account 1234-5678-9012.

1. Nell'account 1111-2222-3333, accedi e apri la console IAM all'indirizzo. AWS Management Console https://console.aws.amazon.com/iam/
2. Crea un nuovo ruolo o visualizza un ruolo esistente che soddisfa i requisiti in. the section called “Requisiti per i ruoli utilizzati per registrare le sedi” Assicurati che il ruolo conceda le autorizzazioni di Amazon S3. `awsexamplebucket1`

3. Apri la console Amazon S3 all'indirizzo https://console.aws.amazon.com/s3/. Accedi con l'account 1234-5678-9012.
4. Nell'elenco dei nomi del bucket, scegli il nome del bucket,. `awsexamplebucket1`
5. Seleziona Autorizzazioni.
6. Nella pagina Autorizzazioni, scegli Bucket Policy.
7. Nell'editor di policy Bucket, incolla la seguente politica. <role-name>Sostituiscilo con il nome del tuo ruolo.

```
{
    "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. Selezionare Salva.
9. Apri la AWS Lake Formation console all'indirizzo https://console.aws.amazon.com/lakeformation/. Accedi all'account 1111-2222-3333 come amministratore del data lake o come utente con autorizzazioni sufficienti per registrare le sedi.
10. Nel riquadro di navigazione, in Amministrazione, scegli Posizioni Data lake.
11. Nella pagina delle sedi dei Data lake, scegli Registra posizione.

12. Nella pagina Registra posizione, per il percorso Amazon S3, inserisci il nome del bucket. `s3://awsexamplebucket1`

> **Note**
>
> Devi digitare il nome del bucket perché i bucket tra account non vengono visualizzati nell'elenco quando scegli Sfoglia.

13. Per il ruolo IAM, scegli il tuo ruolo.
14. Scegli Registra la posizione.

Registrazione di una posizione Amazon S3 crittografata tra più account AWS

AWS Lake Formation si integra con AWS Key Management Service(AWS KMS) per consentirti di configurare più facilmente altri servizi integrati per crittografare e decrittografare i dati nelle sedi Amazon Simple Storage Service (Amazon S3) Simple Storage Service (Amazon S3).

Sono supportate entrambe le chiavi gestite dal cliente. Chiavi gestite da AWS La crittografia/decrittografia lato client non è supportata.

> **Important**
>
> Evita di registrare un bucket Amazon S3 con Requester pay abilitato. Per i bucket registrati con Lake Formation, il ruolo utilizzato per registrare il bucket viene sempre visualizzato come richiedente. Se al bucket si accede da un altro AWS account, al proprietario del bucket viene addebitato l'accesso ai dati se il ruolo appartiene allo stesso account del proprietario del bucket.

Questa sezione spiega come registrare una sede Amazon S3 nelle seguenti circostanze:

- I dati nella posizione Amazon S3 sono crittografati con una chiave KMS creata in. AWS KMS
- La sede Amazon S3 non si trova nello stesso AWS account di. AWS Glue Data Catalog
- La chiave KMS è o non si trova nello stesso AWS account del Data Catalog.

La registrazione di un bucket Amazon S3 AWS KMS crittografato AWS nell'account B utilizzando AWS Identity and Access Management un ruolo (IAM) AWS nell'account A richiede le seguenti autorizzazioni:

- Il ruolo nell'account A deve concedere le autorizzazioni sul bucket dell'account B.
- La politica del bucket nell'account B deve concedere le autorizzazioni di accesso al ruolo nell'account A.
- Se la chiave KMS è nell'account B, la politica chiave deve concedere l'accesso al ruolo nell'account A e il ruolo nell'account A deve concedere le autorizzazioni sulla chiave KMS.

Nella procedura seguente, si crea un ruolo nell' AWS account che contiene il catalogo dati (l'account A nella discussione precedente). Quindi, si utilizza questo ruolo per registrare la posizione. Lake Formation assume questo ruolo quando accede ai dati sottostanti in Amazon S3. Il ruolo assunto dispone delle autorizzazioni richieste sulla chiave KMS. Di conseguenza, non è necessario concedere le autorizzazioni sulla chiave KMS ai responsabili che accedono ai dati sottostanti con lavori ETL o con servizi integrati come. Amazon Athena

> **Important**
>
> Non puoi utilizzare il ruolo collegato al servizio Lake Formation per registrare una sede in un altro account. È invece necessario utilizzare un ruolo definito dall'utente. Il ruolo deve soddisfare i requisiti dithe section called “Requisiti per i ruoli utilizzati per registrare le sedi”. Per ulteriori informazioni sul ruolo collegato al servizio, consulta Autorizzazioni di ruolo collegate al servizio per Lake Formation.

Prima di iniziare

Esamina i requisiti per il ruolo utilizzato per registrare la sede.

Per registrare una posizione Amazon S3 crittografata tra più account AWS

1. Nello stesso AWS account del Data Catalog, accedi AWS Management Console e apri la console IAM all'indirizzohttps://console.aws.amazon.com/iam/.
2. Crea un nuovo ruolo o visualizza un ruolo esistente che soddisfa i requisiti inthe section called “Requisiti per i ruoli utilizzati per registrare le sedi”. Assicurati che il ruolo includa una policy che conceda le autorizzazioni di Amazon S3 sulla location.

3. Se la chiave KMS non si trova nello stesso account del Data Catalog, aggiungi al ruolo una policy in linea che conceda le autorizzazioni richieste sulla chiave KMS. Di seguito è riportata una policy di esempio. Sostituisci <cmk-region>e *< cmk-account-id >* con la regione e il numero di account della chiave KMS. Sostituisci <key-id>con l'ID della chiave.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
        "Effect": "Allow",
        "Action": [
            "kms:Encrypt",
            "kms:Decrypt",
            "kms:ReEncrypt*",
            "kms:GenerateDataKey*",
            "kms:DescribeKey"
         ],
        "Resource": "arn:aws:kms:<cmk-region>:<cmk-account-id>:key/<key-id>"
        }
    ]
}
```

4. Sulla console Amazon S3, aggiungi una bucket policy che conceda le autorizzazioni Amazon S3 richieste per il ruolo. Di seguito è riportato un esempio di policy di bucket. Sostituisci *< catalog-account-id >* con il numero di AWS account del Data Catalog, <role-name>con il nome del tuo ruolo e <bucket-name>con il nome del bucket.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect":"Allow",
            "Principal": {
                "AWS":"arn:aws:iam::<catalog-account-id>:role/<role-name>"
            },
            "Action":"s3:ListBucket",
            "Resource":"arn:aws:s3:::<bucket-name>"
        },
        {
            "Effect":"Allow",
            "Principal": {
                "AWS":"arn:aws:iam::<catalog-account-id>:role/<role-name>"
```

```
            },
            "Action": [
                "s3:DeleteObject",
                "s3:GetObject",
                "s3:PutObject"
            ],
            "Resource":"arn:aws:s3:::<bucket-name>/*"
        }
    ]
}
```

5. In AWS KMS, aggiungi il ruolo come utente della chiave KMS.

   a. Apri la AWS KMS console all'indirizzo https://console.aws.amazon.com/kms. Quindi, accedi come utente amministratore o come utente che può modificare la politica delle chiavi della chiave KMS utilizzata per crittografare la posizione.

   b. Nel riquadro di navigazione, scegli Customer managed keys, quindi scegli il nome della chiave KMS.

   c. Nella pagina dei dettagli della chiave KMS, nella scheda Politica chiave, se la visualizzazione JSON della politica chiave non viene visualizzata, scegli Passa alla visualizzazione delle politiche.

   d. Nella sezione Key policy, scegli Modifica e aggiungi l'Amazon Resource Name (ARN) del ruolo all'`Allow use of the key`oggetto, come mostrato nell'esempio seguente.

> **Note**
>
> Se quell'oggetto manca, aggiungilo con le autorizzazioni mostrate nell'esempio.

```
        ...
        {
            "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": "*"
        },
        ...
```

Per ulteriori informazioni, consulta Consentire agli utenti di altri account di utilizzare una chiave KMS nella Guida per gli AWS Key Management Service sviluppatori.

6. Apri la AWS Lake Formation console all'indirizzo https://console.aws.amazon.com/lakeformation/. Accedi all' AWS account Data Catalog come amministratore del data lake.
7. Nel riquadro di navigazione, in Registra e inserisci, scegli Data lake locations.
8. Scegli Registra posizione.
9. Nella pagina Registra posizione, per il percorso Amazon S3, inserisci il percorso della posizione come. **s3://*<bucket>*/*<prefix>*** Sostituiscilo <bucket>con il nome del bucket e <prefix>con il resto del percorso della posizione.

> **Note**
>
> È necessario digitare il percorso perché i bucket tra account non vengono visualizzati nell'elenco quando si sceglie Sfoglia.

10. Per il ruolo IAM, scegli il ruolo dalla Fase 2.
11. Scegli Registra posizione.

Annullamento della registrazione di una sede Amazon S3

Puoi annullare la registrazione di una sede Amazon Simple Storage Service (Amazon S3) se non desideri più che venga gestita da Lake Formation. L'annullamento della registrazione di una sede non influisce sulle autorizzazioni di localizzazione dei dati di Lake Formation concesse in quella località. È possibile registrare nuovamente una sede che è stata annullata e le autorizzazioni relative alla localizzazione dei dati restano valide. È possibile utilizzare un ruolo diverso per registrare nuovamente la posizione.

Per annullare la registrazione di una posizione (console)

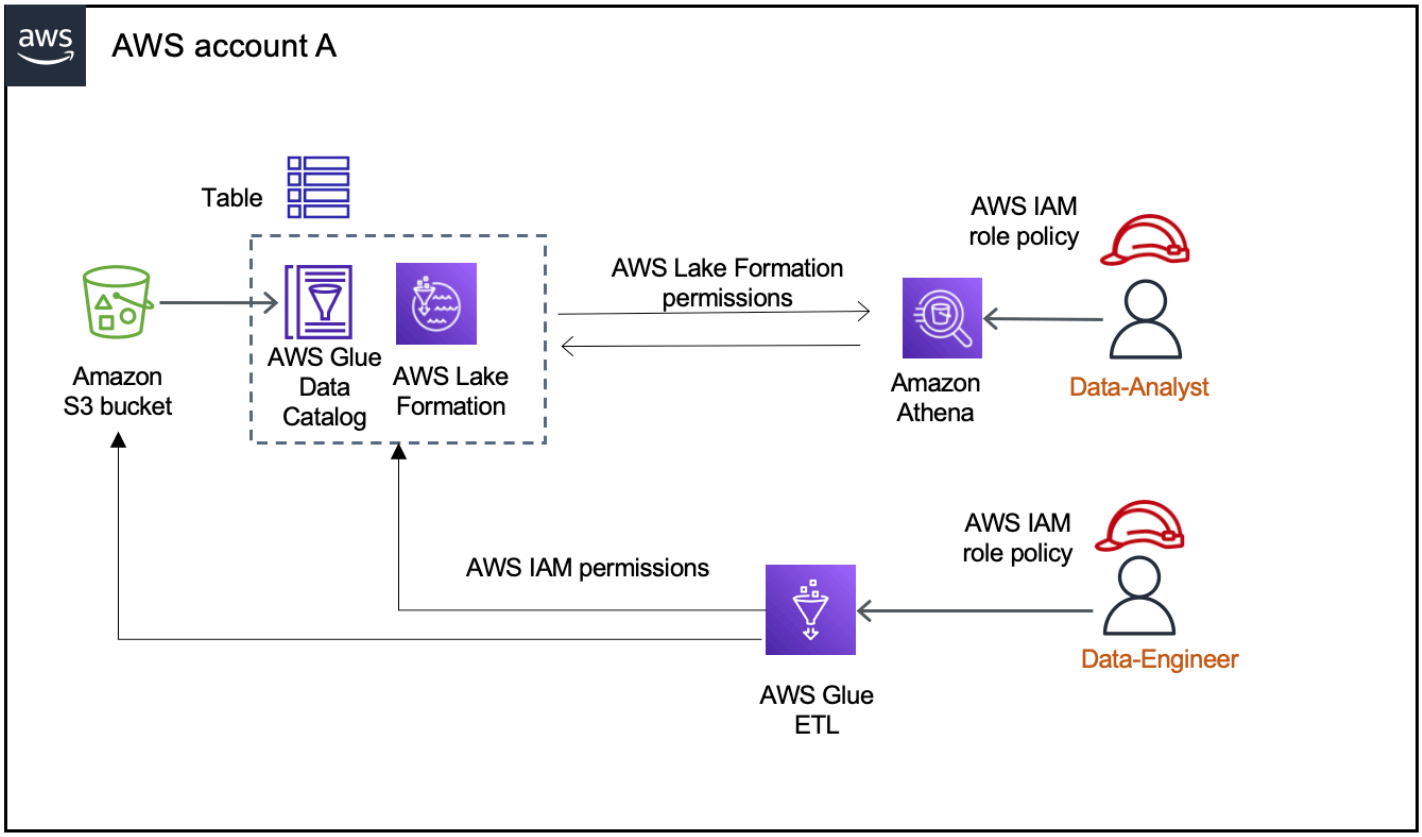
1. Apri la AWS Lake Formation console all'indirizzo https://console.aws.amazon.com/lakeformation/. Accedi come amministratore del data lake o come utente con l'autorizzazione `lakeformation:RegisterResource` IAM.
2. Nel riquadro di navigazione, in Register and Ingest, scegli Data lake locations.
3. Seleziona una posizione e nel menu Azioni scegli Rimuovi.
4. Quando viene richiesta la conferma, scegli Rimuovi.

Modalità di accesso ibrida

AWS Lake Formation la modalità di accesso ibrida supporta due percorsi di autorizzazione per gli stessi database e tabelle. AWS Glue Data Catalog
Nel primo percorso, Lake Formation ti consente di selezionare principali specifici e concedere loro i permessi di Lake Formation per accedere a database e tabelle attivando il consenso. Il secondo percorso consente a tutti gli altri principali di accedere a queste risorse tramite le politiche e le azioni principali IAM predefinite per Amazon AWS Glue S3.

Quando registri una sede Amazon S3 con Lake Formation, hai la possibilità di applicare le autorizzazioni di Lake Formation per tutte le risorse in questa sede o utilizzare la modalità di accesso ibrida. Per impostazione predefinita, la modalità di accesso ibrida applica solo le `CREATE_TABLE` autorizzazioni. `CREATE_PARTITION` `UPDATE_TABLE` Quando una sede Amazon S3 è in modalità ibrida, puoi abilitare le autorizzazioni Lake Formation attivando i principali per database e tabelle in quella posizione.

Pertanto, la modalità di accesso ibrido offre la flessibilità necessaria per abilitare selettivamente Lake Formation per database e tabelle nel tuo Data Catalog per un set specifico di utenti senza interrompere l'accesso per altri utenti o carichi di lavoro esistenti.



Per considerazioni e limitazioni, vedere Considerazioni e limitazioni della modalità di accesso ibrido .

Termini e definizioni

Di seguito sono riportate le definizioni delle risorse di Data Catalog in base alla modalità di impostazione delle autorizzazioni di accesso:

Risorsa Lake Formation

Una risorsa registrata presso Lake Formation. Gli utenti richiedono le autorizzazioni di Lake Formation per accedere alla risorsa.

AWS Glue risorsa

Una risorsa che non è registrata presso Lake Formation. Gli utenti richiedono solo le autorizzazioni IAM per accedere alla risorsa perché dispone di autorizzazioni di `IAMAllowedPrincipals` gruppo. Le autorizzazioni di Lake Formation non vengono applicate.

Per ulteriori informazioni sulle autorizzazioni `IAMAllowedPrincipals` di gruppo, vedere. Autorizzazioni per i metadati

Risorsa ibrida

Una risorsa registrata in modalità di accesso ibrida. In base agli utenti che accedono alla risorsa, la risorsa passa dinamicamente dall'essere una risorsa Lake Formation a una AWS Glue risorsa.

Casi d'uso comuni della modalità di accesso ibrido

Puoi utilizzare la modalità di accesso ibrido per fornire l'accesso in scenari di condivisione dei dati con account singolo e tra account:

Scenari con account singolo

- Convertire una AWS Glue risorsa in una risorsa ibrida: in questo scenario, attualmente non stai utilizzando Lake Formation ma desideri adottare le autorizzazioni Lake Formation per i database e le tabelle di Data Catalog. Quando registri la posizione Amazon S3 in modalità di accesso ibrido, puoi concedere le autorizzazioni di Lake Formation agli utenti che optano per database e tabelle specifici che puntano a quella posizione.
- Convertire una risorsa Lake Formation in una risorsa ibrida: attualmente, utilizzi le autorizzazioni di Lake Formation per controllare l'accesso a un database Data Catalog, ma desideri fornire l'accesso a nuovi principali utilizzando le autorizzazioni IAM per Amazon S3 e AWS Glue senza interrompere le autorizzazioni Lake Formation esistenti.

  Quando aggiorni la registrazione di una posizione dati alla modalità di accesso ibrida, i nuovi responsabili possono accedere al database Data Catalog puntando alla posizione Amazon S3 utilizzando le policy di autorizzazione IAM senza interrompere le autorizzazioni Lake Formation degli utenti esistenti.

  Prima di aggiornare la registrazione della posizione dei dati per abilitare la modalità di accesso ibrida, devi prima attivare i principali che attualmente accedono alla risorsa con le autorizzazioni di Lake Formation.
  Questo serve a prevenire potenziali interruzioni del flusso di lavoro corrente.
  È inoltre necessario concedere al `IAMAllowedPrincipal` gruppo l'`Super`autorizzazione per le tabelle del database.

Scenari di condivisione dei dati tra account

- Condividi AWS Glue risorse utilizzando la modalità di accesso ibrido: in questo scenario, l'account produttore dispone di tabelle in un database che sono attualmente condivise con un account

consumer utilizzando le politiche di autorizzazione IAM per Amazon S3 AWS Glue e le azioni. La posizione dei dati del database non è registrata con Lake Formation.

Prima di registrare la posizione dei dati in modalità di accesso ibrido, è necessario aggiornare le impostazioni della versione dell'account Cross alla versione 4. La versione 4 fornisce le nuove politiche di AWS RAM autorizzazione necessarie per la condivisione tra account quando il `IAMAllowedPrincipal` gruppo dispone dell'`Super`autorizzazione sulla risorsa. Per le risorse con autorizzazioni di `IAMAllowedPrincipal` gruppo, puoi concedere le autorizzazioni di Lake Formation agli account esterni e consentire loro di utilizzare le autorizzazioni Lake Formation. L'amministratore del data lake nell'account del destinatario può concedere le autorizzazioni di Lake Formation ai responsabili dell'account e autorizzarli a far valere le autorizzazioni di Lake Formation.

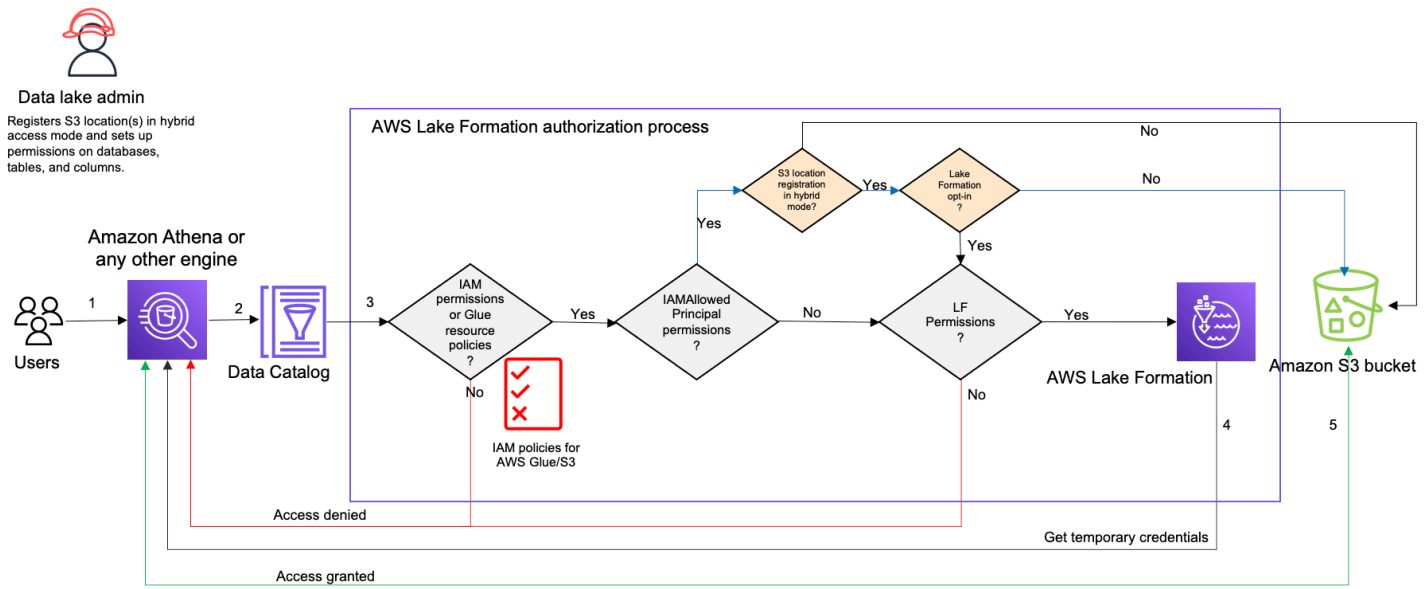
- Condividi le risorse di Lake Formation utilizzando la modalità di accesso ibrida: attualmente, l'account produttore contiene tabelle in un database che vengono condivise con un account consumatore che applica le autorizzazioni di Lake Formation. La posizione dei dati del database è registrata presso Lake Formation.

  In questo caso, puoi aggiornare la registrazione della posizione Amazon S3 alla modalità di accesso ibrida e condividere i dati di Amazon S3 e i metadati di Data Catalog utilizzando le policy dei bucket di Amazon S3 e le politiche delle risorse di Data Catalog con i principali dell'account consumatore. È necessario concedere nuovamente le autorizzazioni esistenti di Lake Formation e attivare i principali prima di aggiornare la registrazione delle sedi Amazon S3. Inoltre, devi concedere al gruppo `Super` l'autorizzazione per le tabelle del database. `IAMAllowedPrincipals`

Argomenti

Come funziona la modalità di accesso ibrida

Il diagramma seguente mostra come funziona l'autorizzazione di Lake Formation in modalità di accesso ibrido quando si interrogano le risorse del Data Catalog.



Prima di accedere ai dati nel data lake, un amministratore del data lake o un utente con autorizzazioni amministrative configura le politiche utente delle singole tabelle di Data Catalog per consentire o negare l'accesso alle tabelle del Data Catalog. Quindi, un principale che dispone delle autorizzazioni per eseguire l'`RegisterResource`operazione registra la posizione Amazon S3 della tabella con Lake Formation in modalità di accesso ibrido. L'amministratore concede le autorizzazioni di Lake Formation a utenti specifici sui database e sulle tabelle del Data Catalog e li autorizza a utilizzare le autorizzazioni Lake Formation per tali database e tabelle in modalità di accesso ibrido.

1. Invia una query: un principale invia una query o uno script ETL utilizzando un servizio integrato come Amazon Athena, Amazon EMR o AWS Glue Amazon Redshift Spectrum.
2. Richiede dati: il motore analitico integrato identifica la tabella richiesta e invia la richiesta di metadati al Data Catalog (,). `GetTable` `GetDatabase`
3. Verifica le autorizzazioni: il Data Catalog verifica le autorizzazioni di accesso del responsabile dell'interrogazione con Lake Formation.
   a. Se alla tabella non sono allegate autorizzazioni di `IAMAllowedPrincipals` gruppo, vengono applicate le autorizzazioni di Lake Formation.
   b. Se il principale ha scelto di utilizzare le autorizzazioni di Lake Formation nella modalità di accesso ibrida e alla tabella sono allegate le autorizzazioni di `IAMAllowedPrincipals` gruppo, le autorizzazioni di Lake Formation vengono applicate. Il motore di query applica i filtri ricevuti da Lake Formation e restituisce i dati all'utente.

c. Se la posizione della tabella non è registrata con Lake Formation e il principale non ha scelto di utilizzare le autorizzazioni di Lake Formation in modalità di accesso ibrido, il Data Catalog verifica se alla tabella sono associate autorizzazioni di `IAMAllowedPrincipals` gruppo. Se questa autorizzazione esiste nella tabella, tutti i principali dell'account ottengono `Super` i `All` permessi sulla tabella.

4. Ottieni credenziali: il Data Catalog controlla e fa sapere al motore se la posizione della tabella è registrata o meno con Lake Formation. Se i dati sottostanti sono registrati con Lake Formation, il motore analitico richiede a Lake Formation le credenziali temporanee per accedere ai dati nel bucket Amazon S3.
5. Ottieni dati: se il responsabile è autorizzato ad accedere ai dati della tabella, Lake Formation fornisce l'accesso temporaneo al motore analitico integrato. Utilizzando l'accesso temporaneo, il motore analitico recupera i dati da Amazon S3 ed esegue i filtri necessari come il filtraggio di colonne, righe o celle. Quando il motore termina l'esecuzione del lavoro, restituisce i risultati all'utente. Questo processo è denominato distribuzione di credenziali. Per ulteriori informazioni, vedereIntegrazione con Lake Formation.
6. Se la posizione dei dati della tabella non è registrata con Lake Formation, la seconda chiamata dal motore di analisi viene effettuata direttamente ad Amazon S3. La policy del bucket Amazon S3 interessata e la politica utente IAM vengono valutate per l'accesso ai dati. Ogni volta che si utilizzano le policy IAM, assicurati di seguire le best practice IAM. Per ulteriori informazioni, consulta le best practice di sicurezza in IAM nella IAM User Guide.

Configurazione della modalità di accesso ibrida: scenari comuni

Come per le autorizzazioni di Lake Formation, in genere sono disponibili due tipi di scenari in cui è possibile utilizzare la modalità di accesso ibrida per gestire l'accesso ai dati: fornire l'accesso ai principali all'interno di uno Account AWS e fornire l'accesso a un principale Account AWS o esterno.

Questa sezione fornisce istruzioni per configurare la modalità di accesso ibrida nei seguenti scenari:

Gestisci le autorizzazioni in modalità di accesso ibrida all'interno di una Account AWS

- Conversione di una AWS Glue risorsa in una risorsa ibrida — Attualmente stai fornendo l'accesso alle tabelle in un database per tutti i principali del tuo account utilizzando le autorizzazioni IAM per Amazon S3 AWS Glue e desideri adottare Lake Formation per gestire le autorizzazioni in modo incrementale.

- Conversione di una risorsa di Lake Formation in una risorsa ibrida — Attualmente stai utilizzando Lake Formation per gestire l'accesso alle tabelle in un database per tutti i principali del tuo account, ma desideri utilizzare Lake Formation solo per i principali specifici. Desideri fornire l'accesso a nuovi principali utilizzando le autorizzazioni IAM per Amazon S3 sullo stesso database AWS Glue e sulle stesse tabelle.

Gestisci le autorizzazioni in modalità di accesso ibrida su s Account AWS

- Condivisione di una AWS Glue risorsa utilizzando la modalità di accesso ibrida— Al momento non stai utilizzando Lake Formation per gestire le autorizzazioni per una tabella, ma desideri applicare le autorizzazioni di Lake Formation per fornire l'accesso ai principali in un altro account.
- Condivisione di una risorsa Lake Formation utilizzando la modalità di accesso ibrida— Stai utilizzando Lake Formation per gestire l'accesso a una tabella ma desideri fornire l'accesso ai principali in un altro account utilizzando le autorizzazioni IAM per Amazon S3 sullo stesso database AWS Glue e sulle stesse tabelle.

Configurazione della modalità di accesso ibrida: passaggi di alto livello

1. Registra la posizione dei dati di Amazon S3 con Lake Formation selezionando la modalità di accesso ibrida.
2. I responsabili devono avere `DATA_LOCATION` l'autorizzazione su una posizione di data lake per creare tabelle o database di Data Catalog che puntano a quella posizione.
3. Imposta l'impostazione della versione per più account sulla versione 4.
4. Concedi autorizzazioni dettagliate a utenti o ruoli IAM specifici su database e tabelle. Allo stesso tempo, assicurati di impostare `Super` o `All` autorizzare il `IAMAllowedPrincipals` gruppo sul database e tutte le tabelle del database o solo alcune di esse.
5. Scegli i principi e le risorse. Gli altri responsabili dell'account possono continuare ad accedere ai database e alle tabelle utilizzando le policy di autorizzazione IAM AWS Glue e le azioni di Amazon S3.
6. Opzionalmente, pulisci le politiche di autorizzazione IAM per Amazon S3 per i principali che hanno scelto di utilizzare le autorizzazioni Lake Formation.

Prerequisiti per la configurazione della modalità di accesso ibrida

Di seguito sono riportati i prerequisiti per l'impostazione della modalità di accesso ibrida:

> **Note**
>
> Consigliamo a un amministratore di Lake Formation di registrare la posizione Amazon S3 in modalità di accesso ibrida e di attivare i principali e le risorse.

1. Concedi l'autorizzazione all'ubicazione dei dati (`DATA_LOCATION_ACCESS`) per creare risorse Data Catalog che puntino alle sedi Amazon S3. Le autorizzazioni di localizzazione dei dati controllano la capacità di creare database e tabelle Data Catalog che puntano a particolari posizioni Amazon S3.
2. Per condividere le risorse di Data Catalog con un altro account in modalità di accesso ibrida (senza rimuovere le autorizzazioni di `IAMAllowedPrincipals` gruppo dalla risorsa), devi aggiornare le impostazioni della versione dell'account Cross alla versione 4. Per aggiornare la versione utilizzando la console Lake Formation, scegli Versione 4 nelle impostazioni della versione dell'account Cross nella pagina delle impostazioni del Data Catalog.

   Puoi anche usare il `put-data-lake-settings` AWS CLI comando per impostare il `CROSS_ACCOUNT_VERSION` parametro sulla versione 4:

   ```
   aws lakeformation put-data-lake-settings --region us-east-1 --data-lake-settings
    file://settings
   {
   "DataLakeAdmins": [
           {
   "DataLakePrincipalIdentifier": "arn:aws:iam::<111122223333>:user/<user-name>"
           }
       ],
       "CreateDatabaseDefaultPermissions": [],
       "CreateTableDefaultPermissions": [],
       "Parameters": {
   "CROSS_ACCOUNT_VERSION": "4"
       }
   }
   ```

3. Per concedere le autorizzazioni su più account in modalità di accesso ibrida, il concedente deve disporre delle autorizzazioni IAM richieste per e i servizi. AWS Glue AWS RAM La policy AWS `AWSLakeFormationCrossAccountManager` gestita concede le autorizzazioni richieste.

Per abilitare la condivisione dei dati tra account in modalità di accesso ibrido, abbiamo aggiornato la policy `AWSLakeFormationCrossAccountManager` gestita aggiungendo due nuove autorizzazioni IAM:

- ram: ListResourceSharePermissions
- ram: AssociateResourceSharePermission

> **Note**
>
> Se non utilizzi la politica AWS gestita per il ruolo di concedente, aggiungi le politiche precedenti alle politiche personalizzate.

Conversione di una AWS Glue risorsa in una risorsa ibrida

Segui questi passaggi per registrare una sede Amazon S3 in modalità di accesso ibrido e inserire nuovi utenti Lake Formation senza interrompere l'accesso ai dati degli utenti esistenti del Data Catalog.

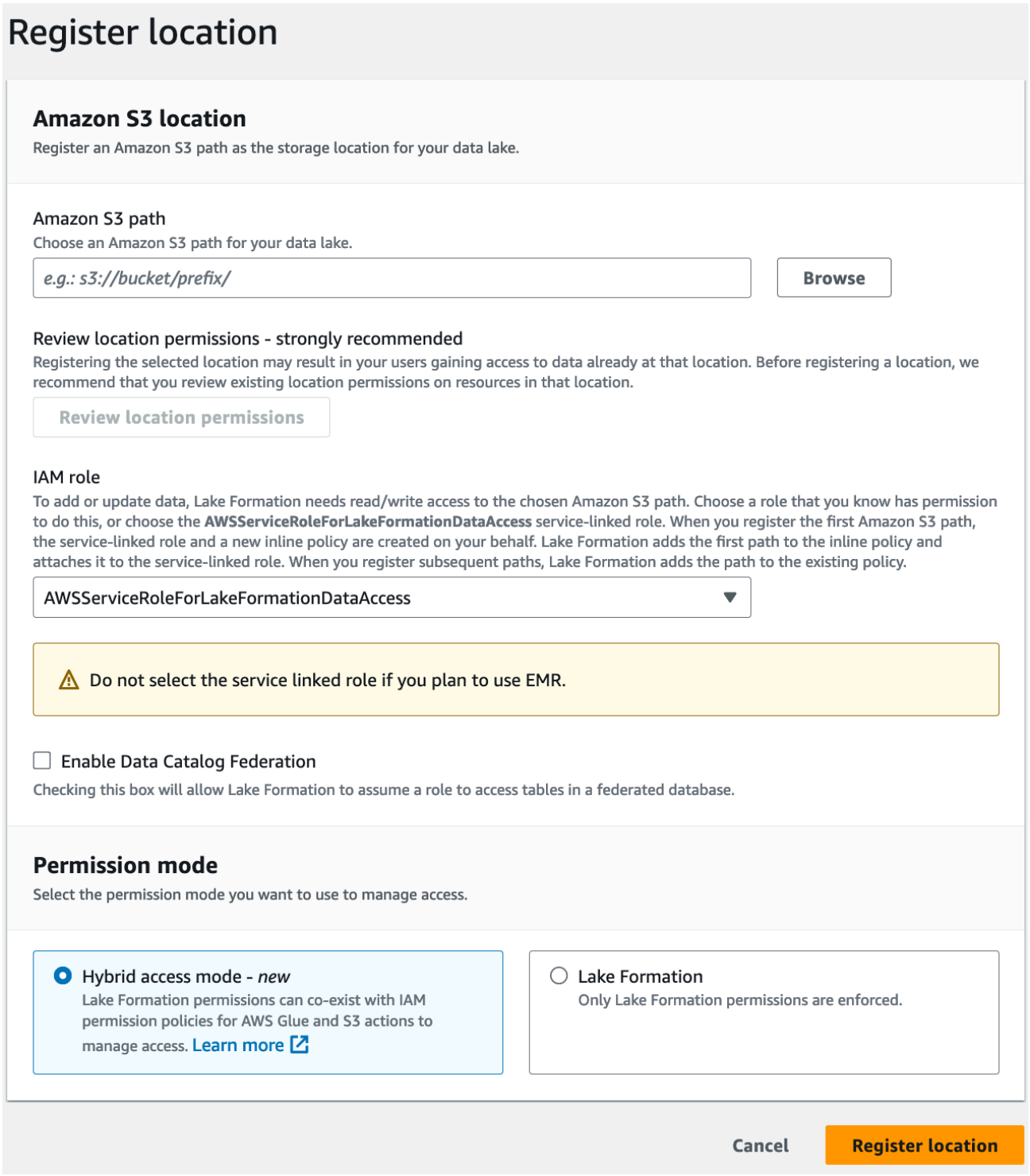
Descrizione dello scenario: la posizione dei dati non è registrata con Lake Formation e l'accesso degli utenti al database e alle tabelle di Data Catalog è determinato dalle politiche di autorizzazione IAM per Amazon AWS Glue S3 e dalle azioni.
Per impostazione predefinita, il `IAMAllowedPrincipals` gruppo dispone di `Super` autorizzazioni su tutte le tabelle del database.

Per abilitare la modalità di accesso ibrido per una posizione dati non registrata con Lake Formation

1. Registra una posizione Amazon S3 abilitando la modalità di accesso ibrida.

   Console

   1. Accedi alla console Lake Formation come amministratore del data lake.
   2. Nel riquadro di navigazione, scegli Posizioni dei data lake in Amministrazione.
   3. Scegli Registra posizione.



4. Nella finestra Registra posizione, scegli il percorso Amazon S3 che desideri registrare con Lake Formation.
5. Per il ruolo IAM, scegli il ruolo `AWSServiceRoleForLakeFormationDataAccess` collegato al servizio (predefinito) o un ruolo IAM personalizzato ruolo che soddisfa i requisiti di. Requisiti per i ruoli utilizzati per registrare le sedi

6. Scegli la modalità di accesso ibrida per applicare politiche di controllo degli accessi granulari di Lake Formation ai principali opt-in e ai database e alle tabelle di Data Catalog che puntano alla posizione registrata.

   Scegli Lake Formation per consentire a Lake Formation di autorizzare le richieste di accesso alla posizione registrata.

7. Scegli Registra posizione.

AWS CLI

Di seguito è riportato un esempio di registrazione di una posizione dati con Lake Formation HybridAccessEnabled con:true/false. Il valore predefinito per il parametro è false. `HybridAccessEnabled` Sostituisci il percorso, il nome del ruolo e l'ID AWS account di Amazon S3 con valori validi.

```
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. Concedi le autorizzazioni e attiva i principali per utilizzare le autorizzazioni di Lake Formation per le risorse in modalità di accesso ibrido

Prima di attivare i principali e le risorse in modalità di accesso ibrido, verifica che la concessione `Super` o `All` le autorizzazioni al `IAMAllowedPrincipals` gruppo esistano sui database e sulle tabelle la cui posizione è registrata con Lake Formation in modalità di accesso ibrida.

> **Note**
>
> Non puoi concedere l'autorizzazione al `IAMAllowedPrincipals` gruppo `All tables` all'interno di un database. È necessario selezionare ogni tabella separatamente dal menu a discesa e concedere le autorizzazioni. Inoltre, quando crei nuove tabelle nel database, puoi scegliere di utilizzarle `Use only IAM access control for new tables in new databases` nelle Impostazioni del catalogo dati. Questa opzione

concede automaticamente l'`Super`autorizzazione al `IAMAllowedPrincipals` gruppo quando si creano nuove tabelle all'interno del database.

Console

1. Nella console Lake Formation, in Data Catalog, scegli Database o Tabelle.
2. Seleziona un database o una tabella dall'elenco e scegli Concedi dal menu Azioni.
3. Scegliete i principali per concedere le autorizzazioni sul database, sulle tabelle e sulle colonne utilizzando il metodo delle risorse denominate o i tag LF.

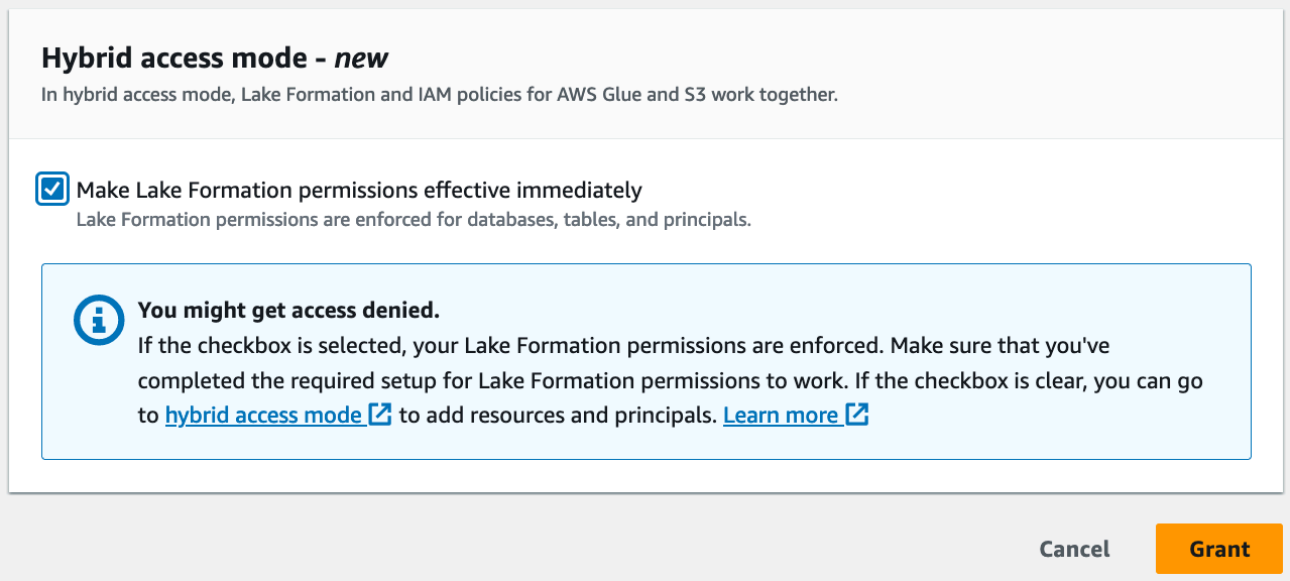
   In alternativa, scegli le autorizzazioni Data lake, seleziona i principali a cui concedere le autorizzazioni dall'elenco e scegli Concedi.

   Per maggiori dettagli sulla concessione delle autorizzazioni per i dati, consulta. Concessione e revoca delle autorizzazioni per le risorse del Data Catalog

   > **Note**
   >
   > Se concedi l'autorizzazione Create table a un principale, devi anche concedere le autorizzazioni per la localizzazione dei dati (`DATA_LOCATION_ACCESS`) al principale. Questa autorizzazione non è necessaria per aggiornare le tabelle. Per ulteriori informazioni, consulta Concessione delle autorizzazioni per la localizzazione dei dati.

4. Quando si utilizza il metodo Named resource per concedere le autorizzazioni, l'opzione per attivare i principali e le risorse è disponibile nella sezione inferiore della pagina Concedi l'autorizzazione ai dati.

   Scegli Rendi immediatamente effettive le autorizzazioni di Lake Formation per abilitare le autorizzazioni di Lake Formation per i committenti e le risorse.



5. Scegli Concessione.

   Quando si attiva il principale A sulla tabella A che punta a una posizione dei dati, consente al principale A di accedere alla posizione di questa tabella utilizzando le autorizzazioni di Lake Formation se la posizione dei dati è registrata in modalità ibrida.

AWS CLI

Di seguito è riportato un esempio di attivazione di un principale e di una tabella in modalità di accesso ibrida. Sostituisci il nome del ruolo, l'id AWS dell'account, il nome del database e il nome della tabella con valori validi.

```
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>"
            }
        }
```

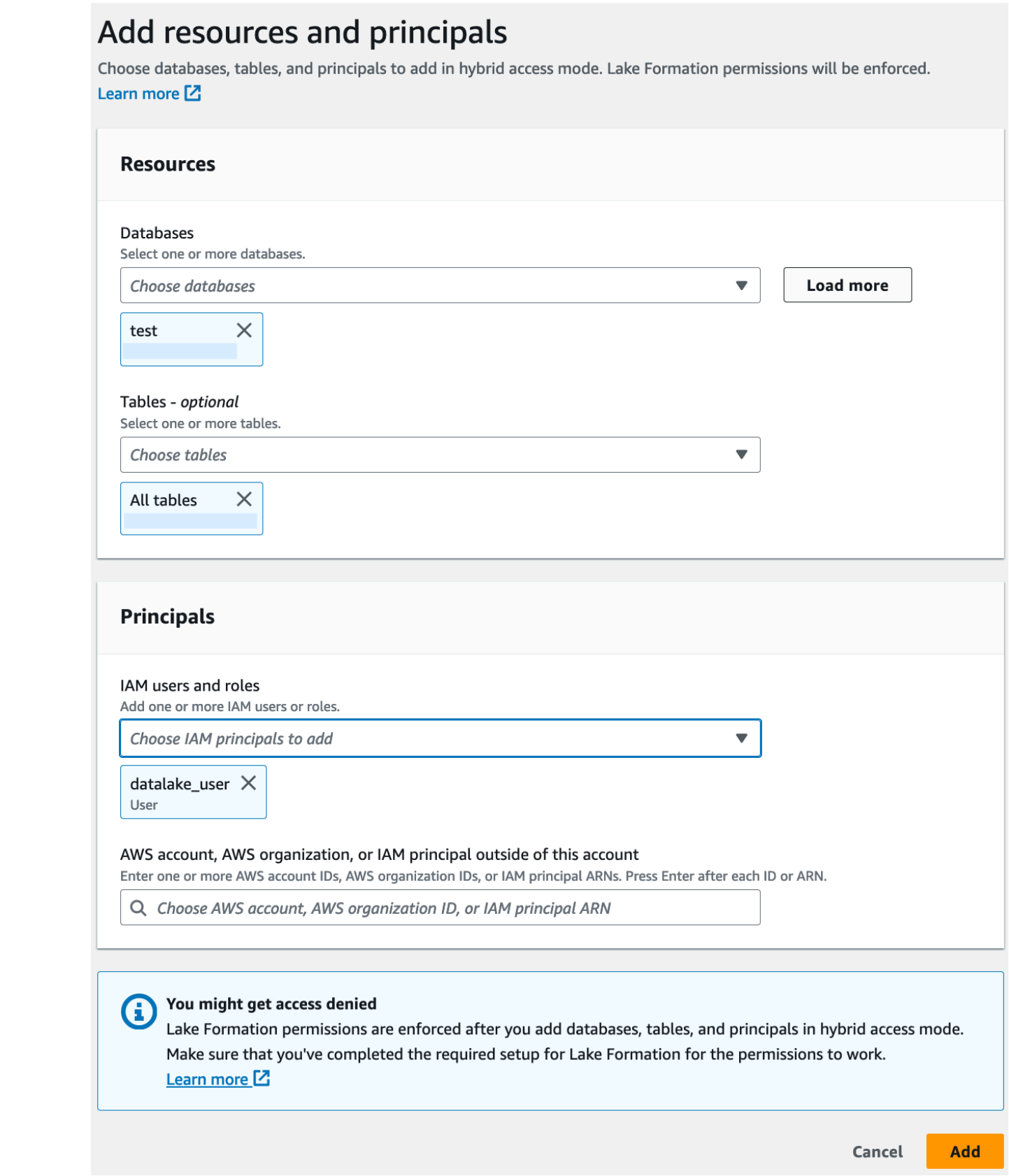
```
    }
```

a. (Optional) Se scegli LF-Tags per concedere le autorizzazioni, puoi attivare i principali per utilizzare i permessi di Lake Formation in un passaggio separato. Puoi farlo scegliendo la modalità di accesso ibrido in Autorizzazioni nella barra di navigazione a sinistra.

b. Nella sezione inferiore della pagina della modalità di accesso ibrida, scegli Aggiungi per aggiungere risorse e principali alla modalità di accesso ibrida.

c. Nella pagina Aggiungi risorse e principali, scegli i database e le tabelle registrati in modalità di accesso ibrido. Scegli i principali a cui attivare l'utilizzo delle autorizzazioni di Lake Formation in modalità di accesso ibrida.

   Puoi scegliere `All tables` all'interno di un database per concedere l'accesso.


Add resources and principals
Choose databases, tables, and principals to add in hybrid access mode. Lake Formation permissions will be enforced.
Learn more
Resources
Databases
Select one or more databases.
Choose databases
Load more
test
Tables - optional
Select one or more tables.
Choose tables
All tables
Principals
IAM users and roles
Add one or more IAM users or roles.
Choose IAM principals to add
datalake_user
User
AWS account, AWS organization, or IAM principal outside of this account
Enter one or more AWS account IDs, AWS organization IDs, or IAM principal ARNs. Press Enter after each ID or ARN.
Choose AWS account, AWS organization ID, or IAM principal ARN
You might get access denied
Lake Formation permissions are enforced after you add databases, tables, and principals in hybrid access mode.
Make sure that you've completed the required setup for Lake Formation for the permissions to work.
Learn more
Cancel
Add

Conversione di una risorsa di Lake Formation in una risorsa ibrida

Nei casi in cui attualmente utilizzi le autorizzazioni di Lake Formation per i database e le tabelle del Data Catalog, puoi modificare le proprietà di registrazione della posizione per abilitare la modalità di accesso ibrida. Ciò consente di fornire ai nuovi principali l'accesso alle stesse risorse utilizzando le politiche di autorizzazione IAM per Amazon S3 AWS Glue e le azioni senza interrompere le autorizzazioni esistenti di Lake Formation.

Descrizione dello scenario: i passaggi seguenti presuppongono che tu abbia una posizione dei dati registrata con Lake Formation e che tu abbia impostato le autorizzazioni per i principali su database, tabelle o colonne che puntano a quella posizione. Se la posizione è stata registrata con un ruolo collegato al servizio, non puoi aggiornare i parametri di posizione e abilitare la modalità di accesso ibrida. Per impostazione predefinita, il `IAMAllowedPrincipals` gruppo dispone di autorizzazioni Super sul database e su tutte le relative tabelle.

> **Important**
>
> Non aggiornate la registrazione di una posizione alla modalità di accesso ibrida senza attivare i principali che accedono ai dati in questa posizione.

Abilitazione della modalità di accesso ibrido per una posizione dati registrata con Lake Formation

1.
> **Warning**
>
> Non consigliamo di convertire una posizione dati gestita da Lake Formation in modalità di accesso ibrida per evitare di interrompere le politiche di autorizzazione di altri utenti o carichi di lavoro esistenti.

   Scegli i presidi esistenti che dispongono dei permessi di Lake Formation.

   1. Elenca e rivedi le autorizzazioni che hai concesso ai principali su database e tabelle. Per ulteriori informazioni, consulta Visualizzazione delle autorizzazioni per database e tabelle in Lake Formation.
   2. Scegli la modalità di accesso ibrido in Autorizzazioni dalla barra di navigazione a sinistra e scegli Aggiungi.

3. Nella pagina Aggiungi principali e risorse, scegli i database e le tabelle dalla posizione dati di Amazon S3 che desideri utilizzare in modalità di accesso ibrido. Scegli i presidi che dispongono già dei permessi di Lake Formation.
4. Scegli Aggiungi per attivare i principali per utilizzare le autorizzazioni di Lake Formation in modalità di accesso ibrida.

2. Aggiorna la registrazione del bucket/prefisso Amazon S3 scegliendo l'opzione della modalità di accesso ibrida.

   Console

   1. Accedi alla console Lake Formation come amministratore del data lake.
   2. Nel riquadro di navigazione, in Register and Ingest, scegli Data lake locations.
   3. Seleziona una posizione e nel menu Azioni scegli Modifica.
   4. Scegli la modalità di accesso ibrida.
   5. Selezionare Salva.
   6. In Data Catalog, seleziona il database o la tabella e `Super` concedi `All` le autorizzazioni al gruppo virtuale chiamato`IAMAllowedPrincipals`.
   7. Verifica che l'accesso degli utenti esistenti di Lake Formation non venga interrotto quando hai aggiornato le proprietà di registrazione della posizione. Accedi alla console Athena come principale di Lake Formation ed esegui una query di esempio su una tabella che punta alla posizione aggiornata.

      Allo stesso modo, verifica l'accesso degli AWS Glue utenti che utilizzano le policy di autorizzazione IAM per accedere al database e alle tabelle.

   AWS CLI

   Di seguito è riportato un esempio di registrazione di una posizione dati con Lake Formation HybridAccessEnabled con:true/false. Il valore predefinito per il parametro è false. `HybridAccessEnabled` Sostituisci il percorso, il nome del ruolo e l'ID AWS account di Amazon S3 con valori validi.

   ```
   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
}
```

Condivisione di una AWS Glue risorsa utilizzando la modalità di accesso ibrida

Condividi i dati con un altro Account AWS o con un responsabile di un altro richiedente Account AWS applicando le autorizzazioni di Lake Formation senza interrompere l'accesso basato su IAM degli utenti di Data Catalog esistenti.

Descrizione dello scenario: l'account produttore dispone di un database Data Catalog il cui accesso è controllato tramite le politiche e AWS Glue le azioni principali di IAM per Amazon S3. La posizione dei dati del database non è registrata con Lake Formation. Per impostazione predefinita, il `IAMAllowedPrincipals` gruppo dispone `Super` delle autorizzazioni per il database e tutte le relative tabelle.

Concessione di autorizzazioni Lake Formation per più account in modalità di accesso ibrida

1. Configurazione dell'account Producer

    1. Accedi alla console Lake Formation utilizzando un ruolo con autorizzazione `lakeformation:PutDataLakeSettings` IAM.
    2. Vai alle impostazioni di Data Catalog e scegli `Version 4` le impostazioni della versione dell'account Cross.

        Se attualmente utilizzi la versione 1 o 2, consulta Aggiornamento delle impostazioni della versione di condivisione dei dati tra account le istruzioni per l'aggiornamento alla versione 3.

        Non sono necessarie modifiche alla politica di autorizzazione per l'aggiornamento dalla versione 3 alla 4.
    3. Registra la posizione Amazon S3 del database o della tabella che intendi condividere in modalità di accesso ibrido.
    4. Verifica che l'`Super`autorizzazione al `IAMAllowedPrincipals` gruppo esista sui database e sulle tabelle di cui hai registrato la posizione dei dati in modalità di accesso ibrida nel passaggio precedente.
    5. Concedi le autorizzazioni di Lake Formation a AWS organizzazioni, unità organizzative (OU) o direttamente con un responsabile IAM in un altro account.

6. Se concedi le autorizzazioni direttamente a un principale IAM, scegli l'indirizzo principale dall'account consumer per applicare le autorizzazioni di Lake Formation in modalità di accesso ibrido abilitando l'opzione Rendi le autorizzazioni di Lake Formation effettive immediatamente.

   Se concedi autorizzazioni per più account a un altro AWS account, quando attivi l'account, le autorizzazioni di Lake Formation vengono applicate solo agli amministratori di quell'account. L'amministratore del data lake dell'account destinatario deve ripartire a cascata le autorizzazioni e attivare i principali dell'account per applicare le autorizzazioni di Lake Formation per le risorse condivise che si trovano in modalità di accesso ibrido.

   Se scegli l'opzione Resources matched by LF-Tags per concedere le autorizzazioni su più account, devi prima completare la fase di concessione delle autorizzazioni. Puoi impostare i principali e le risorse per la modalità di accesso ibrido in un passaggio separato selezionando la modalità di accesso ibrida in Autorizzazioni nella barra di navigazione a sinistra della console Lake Formation. Quindi scegli Aggiungi per aggiungere le risorse e i presidi a cui desideri applicare le autorizzazioni di Lake Formation.

2. Configurazione dell'account consumatore

   1. Accedi alla console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/ come amministratore del data lake.
   2. Vai a https://console.aws.amazon.com/ram e accetta l'invito alla condivisione delle risorse. La scheda Condivisi con me nella AWS RAM console mostra il database e le tabelle condivise con il tuo account.
   3. Crea un collegamento alla risorsa al database e/o alla tabella condivisi in Lake Formation.
   4. Concedi `Describe` l'autorizzazione sul link alla risorsa e l'`Grant on target`autorizzazione (sulla risorsa condivisa originale) ai principali IAM del tuo account (consumatore).
   5. Concedi le autorizzazioni di Lake Formation sul database o sulla tabella condivisa con te ai responsabili del tuo account. Scegli i principi e le risorse per applicare le autorizzazioni di Lake Formation in modalità di accesso ibrida abilitando l'opzione Rendi immediatamente effettive le autorizzazioni di Lake Formation.
   6. Verifica le autorizzazioni Lake Formation del preside eseguendo query Athena di esempio. Testa l'accesso esistente dei tuoi AWS Glue utenti con le policy principali di IAM per Amazon S3 e AWS Glue le azioni.

(Facoltativo) Rimuovi la policy del bucket di Amazon S3 per l'accesso ai dati e le politiche principali di IAM e l'accesso ai dati di Amazon S3 per AWS Glue i principali che hai configurato per utilizzare le autorizzazioni Lake Formation.

Condivisione di una risorsa Lake Formation utilizzando la modalità di accesso ibrida

Consenti ai nuovi utenti di Data Catalog in un account esterno di accedere ai database e alle tabelle di Data Catalog utilizzando policy basate su IAM senza interrompere le autorizzazioni di condivisione tra account esistenti di Lake Formation.

Descrizione dello scenario: l'account produttore dispone di database e tabelle gestiti da Lake Formation condivisi con un account esterno (consumatore) a livello di account o principale IAM. La posizione dei dati del database è registrata presso Lake Formation. Il `IAMAllowedPrincipals` gruppo non dispone `Super` delle autorizzazioni per il database e le relative tabelle.

Garantire l'accesso su più account ai nuovi utenti del Data Catalog tramite policy basate su IAM senza interrompere le autorizzazioni esistenti di Lake Formation

1. Configurazione dell'account Producer

    1. Accedi alla console Lake Formation utilizzando un ruolo che`lakeformation:PutDataLakeSettings`.
    2. In Impostazioni Data Catalog, scegli `Version 4` le impostazioni della versione dell'account Cross.

        Se attualmente utilizzi la versione 1 o 2, consulta Aggiornamento delle impostazioni della versione di condivisione dei dati tra account le istruzioni per l'aggiornamento alla versione 3.

        Non sono necessarie modifiche alla politica di autorizzazione per l'aggiornamento dalla versione 3 alla 4.
    3. Elenca le autorizzazioni che hai concesso ai principali su database e tabelle. Per ulteriori informazioni, consulta Visualizzazione delle autorizzazioni per database e tabelle in Lake Formation.
    4. Rigenera le autorizzazioni esistenti per più account di Lake Formation optando per i principali e le risorse.

> **Note**
>
> Prima di aggiornare la registrazione di una posizione dati alla modalità di accesso ibrida per concedere le autorizzazioni tra più account, devi concedere nuovamente almeno una condivisione di dati tra account per account. Questo passaggio è necessario per aggiornare le autorizzazioni AWS RAM gestite allegate alla condivisione di risorse. AWS RAM
> Nel luglio 2023, Lake Formation ha aggiornato le autorizzazioni AWS RAM gestite utilizzate per la condivisione di database e tabelle:
>
> - `arn:aws:ram::aws:permission/AWSRAMLFEnabledGlueAllTablesReadWriteForDatabase`(politica di condivisione a livello di database)
> - `arn:aws:ram::aws:permission/AWSRAMLFEnabledGlueTableReadWrite`(politica di condivisione a livello di tabella)
>
> Le autorizzazioni per più account concesse prima di luglio 2023 non dispongono di queste autorizzazioni aggiornate. AWS RAM
> Se hai concesso autorizzazioni per più account direttamente ai responsabili, devi concedere nuovamente tali autorizzazioni individualmente ai responsabili. Se salti questo passaggio, i principali che accedono alla risorsa condivisa potrebbero ricevere un errore di combinazione illegale.

5. Vai a https://console.aws.amazon.com/ram.
6. La scheda Condivisi da me nella AWS RAM console mostra i nomi di database e tabelle che hai condiviso con un account o un principale esterno.

   Assicurati che le autorizzazioni allegate alla risorsa condivisa abbiano l'ARN corretto.
7. Verifica che lo stato delle risorse nella AWS RAM condivisione sia attivo. `Associated` Se lo stato è uguale a`Associating`, attendi che entrino `Associated` nello stato. Se lo stato diventa`Failed`, fermati e contatta il team di assistenza di Lake Formation.
8. Scegli la modalità di accesso ibrido in Autorizzazioni dalla barra di navigazione a sinistra e scegli Aggiungi.
9. La pagina Aggiungi principi e risorse mostra i database e/o le tabelle e i principali che hanno accesso. È possibile apportare gli aggiornamenti richiesti aggiungendo o rimuovendo i principali e le risorse.

10. Scegli i principali con autorizzazioni Lake Formation per il database e le tabelle che desideri modificare in modalità di accesso ibrida. Scegli i database e le tabelle.
11. Scegli Aggiungi per attivare i principali per applicare le autorizzazioni di Lake Formation in modalità di accesso ibrida.
12. Concedi `Super` l'autorizzazione al gruppo virtuale `IAMAllowedPrincipals` sul tuo database e sulle tabelle selezionate.
13. Modifica la registrazione della sede Amazon S3 Lake Formation in modalità di accesso ibrida.
14. Concedi le autorizzazioni agli AWS Glue utenti dell'account esterno (consumer) utilizzando le politiche di autorizzazione IAM per le azioni di Amazon AWS Glue S3.

2. Configurazione dell'account consumer

    1. Accedi alla console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/ come amministratore del data lake.
    2. Vai a https://console.aws.amazon.com/ram e accetta l'invito alla condivisione delle risorse. La scheda Risorse condivise con me della AWS RAM pagina mostra i nomi dei database e delle tabelle condivisi con il tuo account.

       Per la AWS RAM condivisione, assicurati che l'autorizzazione allegata contenga l'ARN corretto dell'invito condiviso AWS RAM . Controlla se le risorse della AWS RAM condivisione sono in `Associated` stato. Se lo stato è ugualeAssociating, attendi che diventino `Associated` tali. Se lo stato diventaFailed, fermati e contatta il team di assistenza di Lake Formation.
    3. Crea un collegamento alla risorsa al database e/o alla tabella condivisi in Lake Formation.
    4. Concedi `Describe` l'autorizzazione sul link alla risorsa e l'`Grant on target`autorizzazione (sulla risorsa condivisa originale) ai principali IAM del tuo account (consumatore).
    5. Successivamente, configura le autorizzazioni di Lake Formation per i principali del tuo account nel database o nella tabella condivisa.

       Nella barra di navigazione a sinistra, in Autorizzazioni, scegli la modalità di accesso ibrido.
    6. Scegli Aggiungi nella sezione inferiore della pagina della modalità di accesso ibrido per attivare i principali e il database o la tabella condivisi con te dall'account produttore.
    7. Concedi le autorizzazioni agli AWS Glue utenti del tuo account utilizzando le politiche di autorizzazione IAM per le azioni di Amazon AWS Glue S3.
    8. Verifica le autorizzazioni e AWS Glue le autorizzazioni di Lake Formation degli utenti eseguendo query di esempio separate sulla tabella utilizzando Athena

(Facoltativo) Pulisci le politiche di autorizzazione IAM per Amazon S3 per i principali che si trovano in modalità di accesso ibrido.

Rimozione di principi e risorse dalla modalità di accesso ibrida

Segui questi passaggi per rimuovere database, tabelle e principali dalla modalità di accesso ibrido.

Console

1. Accedi alla console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.
2. In Autorizzazioni, scegli la modalità di accesso ibrida.
3. Nella pagina della modalità di accesso ibrido, seleziona la casella di controllo accanto al nome del database o della tabella e scegli. `Remove`
4. Un messaggio di avviso richiede di confermare l'azione. Scegli Rimuovi.

   Lake Formation non impone più le autorizzazioni per tali risorse e l'accesso a questa risorsa sarà controllato tramite IAM e AWS Glue le autorizzazioni. Ciò potrebbe far sì che l'utente non abbia più accesso a questa risorsa se non dispone delle autorizzazioni IAM appropriate.

AWS CLI

L'esempio seguente mostra come rimuovere risorse dalla modalità di accesso ibrida.

```
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>"
          }
    }
}
```

Visualizzazione dei principali e delle risorse in modalità di accesso ibrida

Segui questi passaggi per visualizzare database, tabelle e principali in modalità di accesso ibrido.

Console

1. Accedi alla console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.
2. In Autorizzazioni, scegli la modalità di accesso ibrida.
3. La pagina della modalità di accesso ibrido mostra le risorse e i principali attualmente in modalità di accesso ibrida.

AWS CLI

L'esempio seguente mostra come elencare tutti i principali e le risorse di opt-in che sono in modalità di accesso ibrido.

```
aws lakeformation list-lake-formation-opt-ins
```

L'esempio seguente mostra come elencare gli opt-in per una specifica coppia principale-risorsa.

```
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>"
          }
```

```
    }
}
```

Risorse aggiuntive

Nel seguente post del blog, ti illustreremo le istruzioni per integrare le autorizzazioni Lake Formation in modalità di accesso ibrida per utenti selezionati mentre il database è già accessibile ad altri utenti tramite le autorizzazioni IAM e Amazon S3. Esamineremo le istruzioni per configurare la modalità di accesso ibrida all'interno di un AWS account e tra due account.

- Presentazione della modalità di accesso AWS Glue Data Catalog ibrida per l'accesso sicuro utilizzando le policy di Lake Formation e IAM e Amazon S3.

Creazione di tabelle e database del catalogo dati

AWS Lake Formationutilizza il AWS Glue Data Catalog per archiviare i metadati relativi ai data lake, alle sorgenti di dati, alle trasformazioni e alle destinazioni. I metadati relativi alle fonti di dati e alle destinazioni sono sotto forma di database e tabelle. Le tabelle memorizzano informazioni sui dati sottostanti, tra cui informazioni sullo schema, sulle partizioni e sulla posizione dei dati. I database sono raccolte di tabelle. Il Data Catalog contiene anche collegamenti a risorse, che sono collegamenti a database e tabelle condivisi in account esterni e vengono utilizzati per l'accesso tra account diversi ai dati nel data lake.

Ogni AWS account dispone di un catalogo dati per AWS regione.

Argomenti

- Creazione di un database
- Creazione di tabelle
- Utilizzo delle visualizzazioni

Creazione di un database

Le tabelle di metadati nel Data Catalog sono archiviate all'interno di database. Puoi creare tutti i database di cui hai bisogno e puoi concedere diverse autorizzazioni Lake Formation su ciascun database.

I database possono avere una proprietà di localizzazione opzionale. Questa sede si trova in genere all'interno di una sede Amazon Simple Storage Service (Amazon S3) registrata presso Lake Formation. Quando specifichi una posizione, i responsabili non necessitano delle autorizzazioni per la localizzazione dei dati per creare tabelle Data Catalog che puntano a posizioni all'interno della posizione del database. Per ulteriori informazioni, consulta Underlying data access control.

Per creare un database utilizzando la console Lake Formation, devi accedere come amministratore del data lake o creatore del database. Un creatore di database è un responsabile a cui è stata concessa l'`CREATE_DATABASE`autorizzazione Lake Formation. Puoi vedere un elenco dei creatori di database nella pagina Ruoli e attività amministrative della console Lake Formation. Per visualizzare questo elenco, è necessario disporre dell'autorizzazione `lakeformation:ListPermissions` IAM e accedere come amministratore del data lake o come creatore di database con l'opzione di concessione sull'`CREATE_DATABASE`autorizzazione.

Per creare un database:

1. Apri la AWS Lake Formation console all'indirizzo https://console.aws.amazon.com/lakeformation/ e accedi come amministratore del data lake o creatore di database.
2. Nel riquadro di navigazione, in Catalogo dati, scegli Database.
3. Scegliere Crea database.
4. Nella finestra di dialogo Crea database, inserisci un nome di database, una posizione opzionale e una descrizione facoltativa.
5. Facoltativamente, seleziona Usa solo il controllo di accesso IAM per le nuove tabelle in questo database.

   Per ulteriori informazioni su questa opzione, consulta the section called “Modifica delle impostazioni predefinite per il data lake”.
6. Scegliere Crea database.

Creazione di tabelle

AWS Lake Formationle tabelle di metadati contengono informazioni sui dati nel data lake, tra cui informazioni sullo schema, sulle partizioni e sulla posizione dei dati. Queste tabelle sono archiviate nel AWS Glue Data Catalog. Li usi per accedere ai dati sottostanti nel data lake e gestirli con le autorizzazioni di Lake Formation. Le tabelle vengono archiviate all'interno dei database del Data Catalog.

Esistono diversi modi per creare tabelle Data Catalog:

- Esegui un crawler. AWS Glue Vedi Definizione dei crawler nella Guida per gli sviluppatori. AWS Glue
- Crea ed esegui un flusso di lavoro. Per informazioni, consulta the section called “Importazione di dati tramite flusso di lavoro”.
- Crea una tabella manualmente utilizzando la console di Lake Formation, AWS Glue l'API o AWS Command Line Interface (AWS CLI).
- Crea una tabella utilizzandoAmazon Athena.
- Crea un collegamento di risorsa a una tabella in un account esterno. Per informazioni, consulta the section called “Creazione di collegamenti alle risorse”.

Creazione di tabelle Apache Iceberg

AWS Lake Formationsupporta la creazione di tabelle Apache Iceberg che utilizzano il formato di dati Apache Parquet AWS Glue Data Catalog con dati che risiedono in Amazon S3. Una tabella nel Data Catalog è la definizione di metadati che rappresenta i dati in un data store. Per impostazione predefinita, Lake Formation crea tabelle Iceberg v2. Per la differenza tra le tabelle v1 e v2, consulta Modifiche al tipo di formato nella documentazione di Apache Iceberg.

Apache Iceberg è un formato a tabella aperta per set di dati analitici di dimensioni molto grandi. Iceberg consente di modificare facilmente lo schema, noto anche come evoluzione dello schema, il che significa che gli utenti possono aggiungere, rinominare o rimuovere colonne da una tabella di dati senza interrompere i dati sottostanti. Iceberg fornisce anche supporto per il controllo delle versioni dei dati, che consente agli utenti di tenere traccia delle modifiche ai dati nel tempo. Ciò abilita la funzionalità Time Travel, che consente agli utenti di accedere e interrogare le versioni storiche dei dati e analizzare le modifiche ai dati tra aggiornamenti ed eliminazioni.

Puoi utilizzare la console Lake Formation o l'`CreateTable`operazione nell'AWS GlueAPI per creare una tabella Iceberg nel Data Catalog. Per ulteriori informazioni, vedere CreateTable action (Python: create_table).

Quando crei una tabella Iceberg nel Data Catalog, devi specificare il formato della tabella e il percorso del file dei metadati in Amazon S3 per poter eseguire letture e scritture.

Puoi usare Lake Formation per proteggere la tua tabella Iceberg utilizzando autorizzazioni di controllo degli accessi granulari quando registri la posizione dati di Amazon S3 con. AWS Lake Formation Per i dati di origine in Amazon S3 e i metadati che non sono registrati con Lake Formation, l'accesso

è determinato dalle policy di autorizzazione IAM per Amazon S3 e dalle operazioni AWS Glue. Per ulteriori informazioni, consulta Gestione delle autorizzazioni di Lake Formation.

> **Note**
>
> Data Catalog non supporta la creazione di partizioni e l'aggiunta di proprietà delle tabelle Iceberg.

Argomenti

Prerequisiti

Per creare tabelle Iceberg nel Data Catalog e configurare le autorizzazioni di accesso ai dati di Lake Formation, devi soddisfare i seguenti requisiti:

1. Autorizzazioni necessarie per creare tabelle Iceberg senza i dati registrati con Lake Formation.

   Oltre alle autorizzazioni necessarie per creare una tabella nel Data Catalog, il creatore della tabella richiede le seguenti autorizzazioni:

   - `s3:PutObject`sulla risorsa arn:aws:s3::: {bucketName}
   - `s3:GetObject`sulla risorsa arn:aws:s3::: {bucketName}
   - `s3:DeleteObject`sulla risorsa arn:aws:s3::: {bucketName}
2. Autorizzazioni necessarie per creare tabelle Iceberg con dati registrati con Lake Formation:

   Per utilizzare Lake Formation per gestire e proteggere i dati nel tuo data lake, registra la tua posizione Amazon S3 che contiene i dati per le tabelle con Lake Formation. In questo modo Lake Formation può fornire credenziali a servizi di AWS analisi come Athena, Redshift Spectrum e Amazon EMR per accedere ai dati. Per ulteriori informazioni sulla registrazione di una sede Amazon S3, consulta. Aggiungere una posizione Amazon S3 al tuo data lake

   Un preside che legge e scrive i dati sottostanti registrati con Lake Formation richiede le seguenti autorizzazioni:

   - `lakeformation:GetDataAccess`

- `DATA_LOCATION_ACCESS`

  Un responsabile che dispone delle autorizzazioni di localizzazione dei dati su una sede dispone anche delle autorizzazioni di localizzazione su tutte le sedi dei figli.

  Per ulteriori informazioni sulle autorizzazioni per la localizzazione dei dati, vedere. Controllo sottostante dell'accesso ai dati

Per abilitare la compattazione, il servizio deve assumere un ruolo IAM con le autorizzazioni per aggiornare le tabelle nel Data Catalog. Per maggiori dettagli, consulta Prerequisiti per l'ottimizzazione delle tabelle .

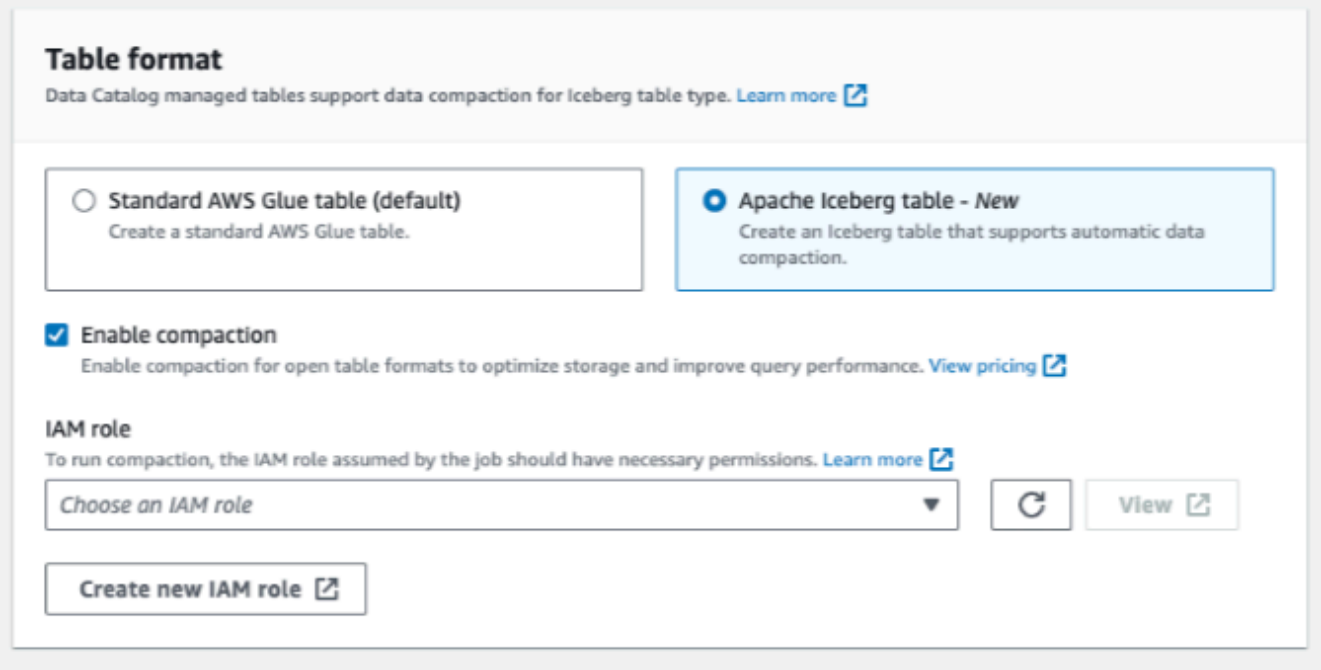
Creazione di una tabella Iceberg

Puoi creare tabelle Iceberg v1 e v2 utilizzando la console Lake Formation o AWS Command Line Interface come documentato in questa pagina. Puoi anche creare tabelle Iceberg usando la console o. AWS Glue Crawler di AWS Glue Per ulteriori informazioni, consulta Data Catalog and Crawlers nella Developer Guide. AWS Glue

Per creare una tabella Iceberg

Console

1. Accedi a e apri AWS Management Console la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.
2. In Data Catalog, scegliete Tabelle e utilizzate il pulsante Crea tabella per specificare i seguenti attributi:
   - Nome tabella: immettete un nome per la tabella. Se utilizzi Athena per accedere alle tabelle, utilizza questi suggerimenti di denominazione nella Amazon Athena User Guide.
   - Database: scegli un database esistente o creane uno nuovo.
   - Descrizione: la descrizione della tabella. Puoi scrivere una descrizione per aiutarti a comprendere i contenuti della tabella.
   - Formato tabella: per il formato tabella, scegli Apache Iceberg.



- Abilita la compattazione: scegli Abilita compattazione per compattare piccoli oggetti Amazon S3 nella tabella in oggetti più grandi.
- Ruolo IAM: per eseguire la compattazione, il servizio assume un ruolo IAM per tuo conto. Puoi scegliere un ruolo IAM utilizzando il menu a discesa. Assicurati che il ruolo disponga delle autorizzazioni necessarie per abilitare la compattazione.

  Per ulteriori informazioni sulle autorizzazioni richieste, consulta. Prerequisiti per l'ottimizzazione delle tabelle
- Posizione: specifica il percorso della cartella in Amazon S3 che memorizza la tabella dei metadati. Iceberg necessita di un file di metadati e di una posizione nel Data Catalog per poter eseguire letture e scritture.
- Schema: scegli Aggiungi colonne per aggiungere colonne e tipi di dati delle colonne. Hai la possibilità di creare una tabella vuota e aggiornare lo schema in un secondo momento. Data Catalog supporta i tipi di dati Hive. Per ulteriori informazioni, consulta Tipi di dati Hive.

  Iceberg consente di evolvere lo schema e la partizione dopo aver creato la tabella. Puoi usare le query Athena per aggiornare lo schema della tabella e le query Spark per aggiornare le partizioni.

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/"
            }
        }'
```

Ottimizzazione delle tabelle Iceberg

I data lake Amazon S3 che utilizzano formati di tabelle aperte, come Apache Iceberg, archiviano i dati come oggetti Amazon S3. La presenza di migliaia di piccoli oggetti Amazon S3 in una tabella di data lake aumenta il sovraccarico dei metadati sulle tabelle Iceberg e influisce sulle prestazioni di lettura. Per migliorare le prestazioni di lettura tramite servizi di analisi AWS come Amazon Athena e Amazon EMR e processi AWS Glue ETL, AWS Glue Data Catalog offre la compattazione gestita (un processo che compatta piccoli oggetti Amazon S3 in oggetti più grandi) per le tabelle Iceberg in Catalogo dati. Puoi utilizzare la console, AWS Glue la console o l'AWSAPI di Lake Formation per abilitare o disabilitare la compattazione per le singole tabelle Iceberg presenti nel Data Catalog. AWS CLI

L'ottimizzatore delle tabelle monitora costantemente le partizioni delle tabelle e avvia il processo di compattazione quando viene superata la soglia per il numero di file e le dimensioni dei file. Nel Catalogo dati, il valore di soglia predefinito per avviare la compattazione è impostato su 384 MB, mentre nella libreria Iceberg la soglia per la compattazione è di circa 75% rispetto alla dimensione del file di destinazione. Catalogo dati esegue la compattazione senza interferire con le query simultanee. Catalogo dati supporta la compattazione dei dati solo per le tabelle in formato Parquet.

Per i tipi di dati, i formati di compressione e le limitazioni supportati, consulta. Formati e limitazioni supportati per la compattazione gestita dei dati

Argomenti

Prerequisiti per l'ottimizzazione delle tabelle

L'ottimizzatore di tabelle assume le autorizzazioni del ruolo (IAM) AWS Identity and Access Management specificate quando si abilita la compattazione per una tabella. Il ruolo IAM deve disporre delle autorizzazioni per leggere i dati e aggiornare i metadati nel Catalogo dati. Puoi creare un ruolo IAM e collegare le seguenti policy in linea:

- Aggiungi la seguente policy in linea che concede le autorizzazioni di lettura/scrittura di Amazon S3 sulla posizione per i dati non registrati con Lake Formation. Questa politica include anche le autorizzazioni per aggiornare la tabella nel Catalogo dati e consentire a AWS Glue di aggiungere log nei log Amazon CloudWatch e la pubblicazione di parametri. Per i dati di origine in Amazon S3 che non sono registrati con Lake Formation, l'accesso è determinato dalle policy di autorizzazione IAM per Amazon S3 e dalle operazioni AWS Glue.

  Nelle seguenti policy in linea, sostituisci `bucket-name` con il nome del bucket Amazon S3, `aws-account-id` e `region` con un numero di account AWS valido e una regione del Catalogo dati, `database_name` con il nome del database e `table_name` con il nome della tabella.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "s3:PutObject",
                "s3:GetObject",
                "s3:DeleteObject"
```

```
            ],
            "Resource": [
                "arn:aws:s3:::<bucket-name>/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::<bucket-name>"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "glue:UpdateTable",
                "glue:GetTable"
            ],
            "Resource": [
                "arn:aws:glue:<region>:<aws-account-id>:table/<database-name>/<table-
name>",
                "arn:aws:glue:<region>:<aws-account-id>:database/<database-name>",
                "arn:aws:glue:<region>:<aws-account-id>:catalog"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "logs:CreateLogGroup",
                "logs:CreateLogStream",
                "logs:PutLogEvents"
            ],
            "Resource": "arn:aws:logs:<region>:<aws-account-id>:log-group:/aws-glue/
iceberg-compaction/logs:*"
        }
    ]
}
```

- Utilizza la seguente policy per abilitare la compattazione dei dati registrati con Lake Formation.

Se al ruolo di compattazione non sono concesse le autorizzazioni di `IAM_ALLOWED_PRINCIPALS` gruppo sulla tabella, il ruolo richiede Lake Formation `ALTER` `INSERT` e `DELETE` le autorizzazioni sulla tabella. `DESCRIBE`

Per ulteriori informazioni sulla registrazione di un bucket Amazon S3 con Lake Formation, consulta. [Aggiungere una posizione Amazon S3 al tuo data lake]

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

        "Resource": [
            "arn:aws:glue:<region>:<aws-account-
id>:table/<databaseName>/<tableName>",
            "arn:aws:glue:<region>:<aws-account-id>:database/<database-name>",
            "arn:aws:glue:<region>:<aws-account-id>:catalog"
        ]
    },
    {
        "Effect": "Allow",
        "Action": [
            "logs:CreateLogGroup",
            "logs:CreateLogStream",
            "logs:PutLogEvents"
        ],
        "Resource": "arn:aws:logs:<region>:<aws-account-id>:log-group:/aws-glue/
iceberg-compaction/logs:*"
        }
    ]
```

```
 }
```

- (Facoltativo) Per compattare le tabelle Iceberg con i dati nei bucket Amazon S3 crittografati utilizzando la Crittografia lato server, il ruolo di compattazione richiede le autorizzazioni per decrittografare gli oggetti Amazon S3 e per generare una nuova chiave dati per scrivere oggetti nei bucket crittografati. Aggiungere la policy seguente alla chiave AWS KMS desiderata. Supportiamo solo la crittografia a livello di bucket.

```
{
    "Effect": "Allow",
    "Principal": {
        "AWS": "arn:aws:iam::<aws-account-id>:role/<compaction-role-name>"
    },
    "Action": [
        "kms:Decrypt",
        "kms:GenerateDataKey"
    ],
    "Resource": "*"
}
```

- (Facoltativo) Per la posizione dei dati registrati con Lake Formation, il ruolo utilizzato per registrare la posizione richiede le autorizzazioni per decrittografare gli oggetti Amazon S3 e generare una nuova chiave dati per scrivere oggetti nei bucket crittografati. Per ulteriori informazioni, consulta Registrazione di una posizione Amazon S3 crittografata.
- (Facoltativo) Se la AWS KMS chiave è archiviata in un altro AWS account, devi includere le seguenti autorizzazioni per il ruolo di compattazione.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "kms:Decrypt",
                "kms:GenerateDataKey"
            ],
            "Resource": ["arn:aws:kms:<REGION>:<KEY_OWNER_ACCOUNT_ID>:key/<KEY_ID>" ]
        }
    ]
}
```

- Il ruolo utilizzato per eseguire la compattazione deve disporre dell'autorizzazione `iam:PassRole` relativa al ruolo.

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

- Aggiungi la seguente policy di attendibilità al ruolo per il servizio AWS Glue per assumere il ruolo IAM ed eseguire il processo di compattazione.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "",
            "Effect": "Allow",
            "Principal": {
                "Service": "glue.amazonaws.com"
            },
            "Action": "sts:AssumeRole"
        }
    ]
 }
```

Abilitazione della compattazione

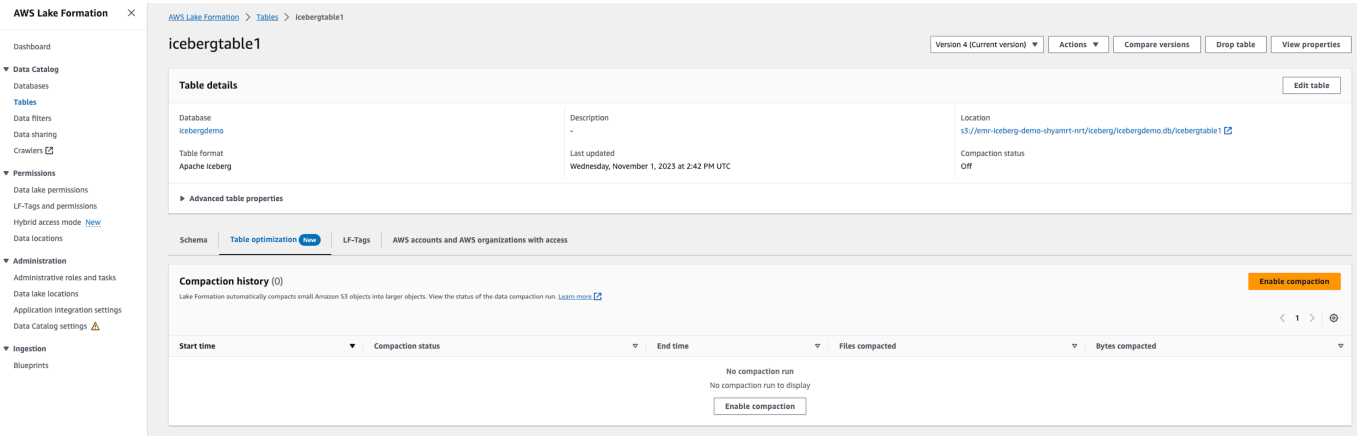
Puoi utilizzare la console, AWS Glue la console o l'AWSAPI di Lake Formation per abilitare la compattazione delle tabelle Apache Iceberg nel Data Catalog. AWS CLI Per le nuove tabelle, puoi

scegliere Apache Iceberg come formato di tabella e abilitare la compattazione quando crei la tabella. La compattazione è disabilitata per impostazione predefinita per le nuove tabelle.

Console

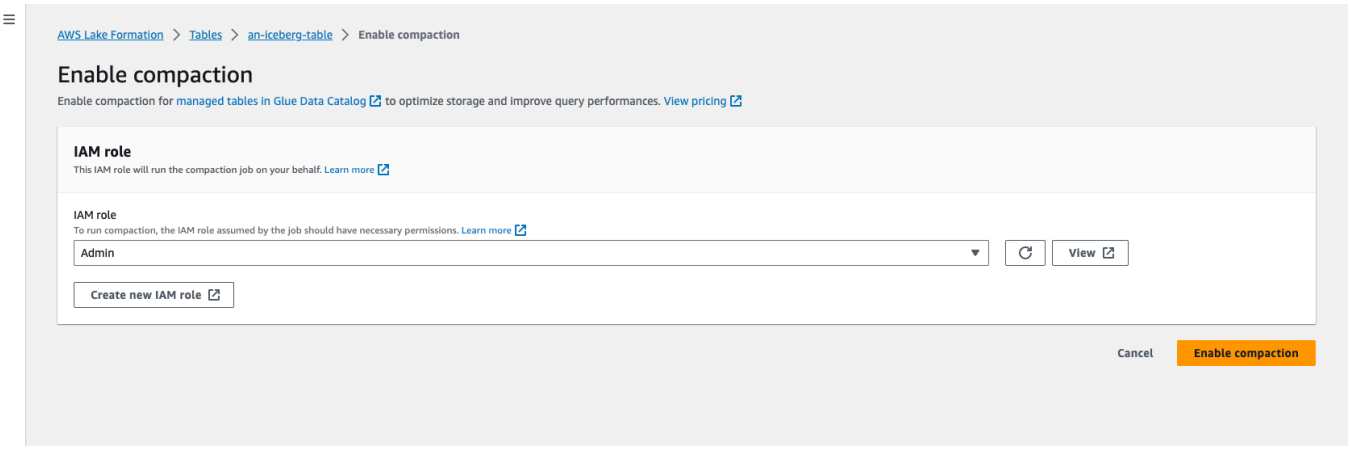
Per abilitare la compattazione

1. Apri la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/ e accedi come amministratore del data lake, creatore della tabella o utente a cui sono state `glue:UpdateTable` concesse le `lakeformation:GetDataAccess` autorizzazioni and sulla tabella.
2. Nel pannello di navigazione, in Catalogo dati, seleziona Tabelle.
3. Nella pagina Tabelle, scegli una tabella in formato tabella aperta per la quale desideri abilitare la compattazione, quindi nel menu Azioni, scegli Abilita compattazione.
4. Puoi anche abilitare la compattazione selezionando la tabella e aprendo la pagina dei Dettagli della tabella. Scegli la scheda Ottimizzazione della tabella nella sezione inferiore della pagina e scegli Abilita compattazione.



5. Successivamente, seleziona un ruolo IAM esistente dal menu a discesa con le autorizzazioni mostrate nella sezione Prerequisiti per l'ottimizzazione delle tabelle.

   Quando scegli l'opzione Crea un nuovo ruolo IAM, il servizio crea un ruolo personalizzato con le autorizzazioni necessarie per eseguire la compattazione.



Segui la procedura riportata di seguito per aggiornare un ruolo IAM esistente:

a. Per aggiornare la politica di autorizzazione per il ruolo IAM, nella console IAM, vai al ruolo IAM utilizzato per eseguire la compattazione.

b. Nella sezione Autorizzazioni, scegli Aggiungi policy bucket. Nella finestra del browser appena aperta, crea una nuova policy da utilizzare con il tuo ruolo.

c. Nella pagina Crea policy, scegli la scheda JSON. Copia il codice JSON mostrato nei Prerequisiti nel campo dell'editor delle politiche.

AWS CLI

L'esempio seguente mostra come abilitare la compattazione. Sostituisci l'ID dell'account con un ID dell'account AWS valido. Sostituisci il nome del database e della tabella con quello effettivo della tabella Iceberg e del database. Sostituisci `roleArn` con il nome della risorsa (ARN) AWS del ruolo IAM e il nome del ruolo IAM che dispone delle autorizzazioni necessarie per eseguire la compattazione.

```
aws glue create-table-optimizer \
  --catalog-id 123456789012 \
  --database-name iceberg_db \
  --table-name iceberg_table \
  --table-optimizer-configuration
 '{"roleArn":"arn:aws:iam::123456789012:role/compaction_role", "enabled":'true'}' \
  --type compaction
```

AWS API

Chiama l'operazione `CreateTableOptimizer` per abilitare la compattazione di una tabella.

Dopo aver abilitato la compattazione, la scheda di Ottimizzazione della tabella mostra i seguenti dettagli di compattazione, dopo circa 15-20 minuti:

Ora di inizio

Il momento in cui è iniziato il processo di compattazione all'interno di Lake Formation. Il valore è un timestamp in formato UTC.

Ora di fine

L'ora in cui il processo di compattazione è terminato in Data Catalog. Il valore è un timestamp in formato UTC.

Stato

Lo stato del ciclo di compattazione. I valori sono esito positivo o negativo.

File compattati

Numero totale di file compattati.

Byte compattati

Numero totale di byte compattati.

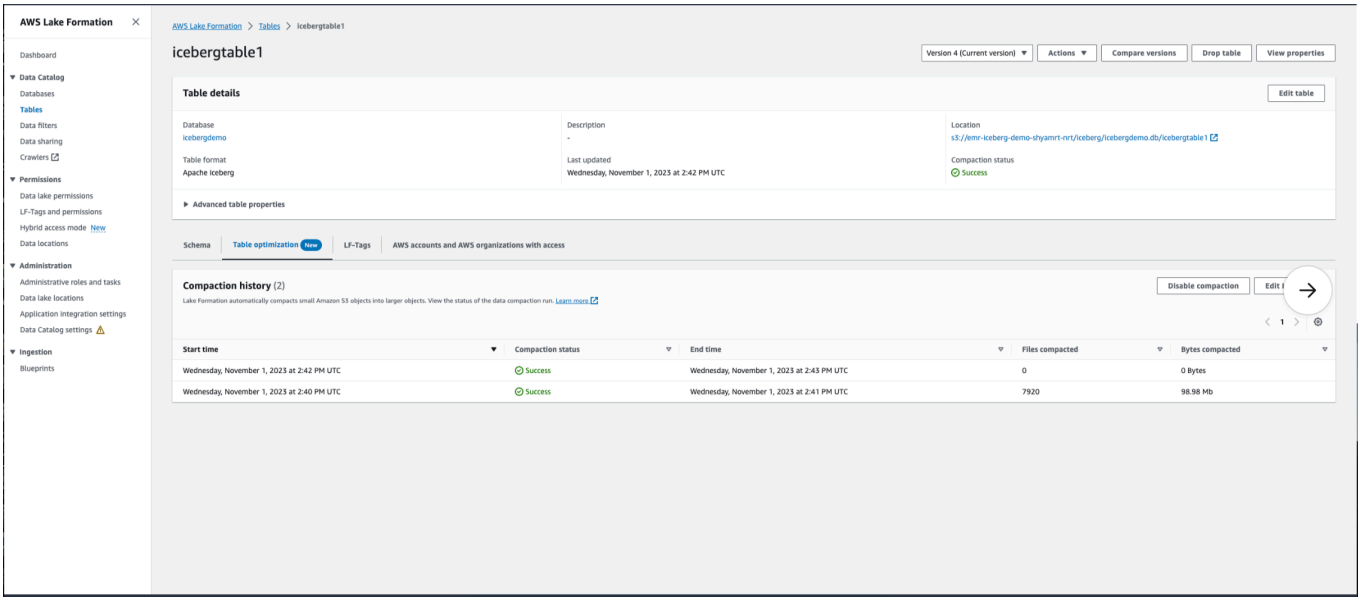
Disabilitazione della compattazione

È possibile disabilitare la compattazione automatica per una particolare tabella Apache Iceberg utilizzando la console AWS Glue oppure AWS CLI.

Console

1. Scegli Catalogo dati e poi Tabelle. Dall'elenco delle tabelle, scegli la tabella in formato tabella aperta di cui desideri disabilitare la compattazione.
2. Puoi scegliere una tabella Iceberg e scegliere Disabilita compattazione in Azioni.

   Puoi anche disabilitare la compattazione per la tabella scegliendo Disabilita compattazione nella sezione inferiore della pagina dei Dettagli delle tabelle.



3. Scegli Disabilita la compattazione nel messaggio di conferma. È possibile abilitare nuovamente la compattazione in un secondo momento.

   Dopo la conferma, la compattazione viene disabilitata e il relativo stato torna a `Off`.

AWS CLI

Nell'esempio seguente, sostituisci l'ID account con un ID account AWS valido. Sostituisci il nome del database e della tabella con quello effettivo della tabella Iceberg e del database. Sostituisci `roleArn` con il nome della risorsa (ARN) AWS del ruolo IAM e il nome effettivo del ruolo IAM che dispone delle autorizzazioni necessarie per eseguire la compattazione.

```
aws glue update-table-optimizer \
  --catalog-id 123456789012 \
  --database-name iceberg_db \
  --table-name iceberg_table \
  --table-optimizer-configuration
 '{"roleArn":"arn:aws:iam::123456789012:role/compaction_role", "enabled":'false'}'\
  --type compaction
```

AWS API

UpdateTableOptimizer Operazione di chiamata per disabilitare la compattazione per una tabella specifica.

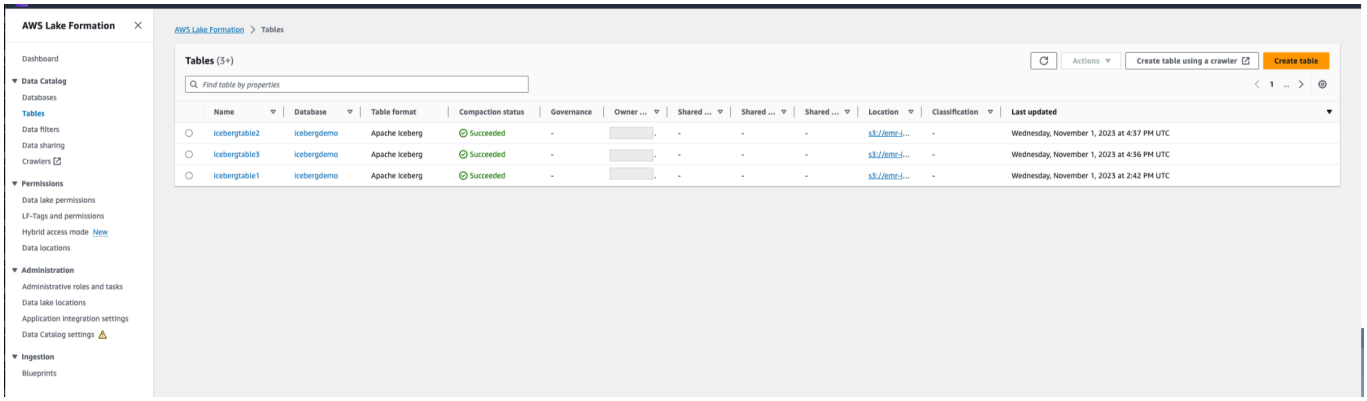
Visualizzazione dei dettagli della compattazione

È possibile visualizzare lo stato di compattazione per Apache Iceberg nella console di Lake Formation o utilizzando AWS CLI AWS le operazioni API.

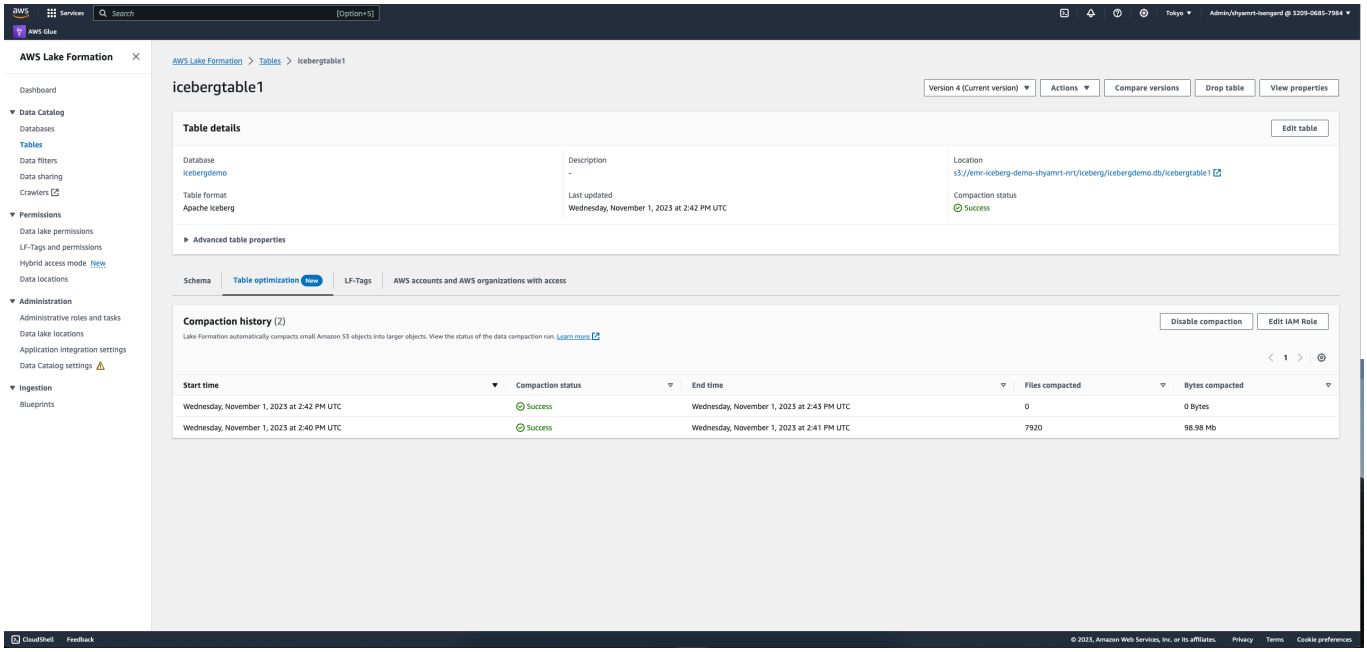
Console

Per visualizzare lo stato di compattazione delle tabelle Iceberg (console)

- Puoi visualizzare lo stato di compattazione delle tabelle Iceberg sulla console Lake Formation selezionando Tabelle in Data Catalog. Il campo Stato di compattazione mostra lo stato dell'esecuzione di compattazione. È possibile visualizzare il formato della tabella e lo stato di compattazione utilizzando le preferenze della tabella.



- Per visualizzare la cronologia delle esecuzioni di compattazione per una tabella specifica, scegli Tabelle in AWS Glue Data Catalog e scegli una tabella per visualizzarne i dettagli. La scheda Ottimizzazione della tabella ne mostra la cronologia di compattazione.



AWS CLI

È possibile visualizzare i dettagli della compattazione utilizzando AWS CLI.

Negli esempi seguenti, sostituite l'ID account con un ID account AWS valido, il nome del database e della tabella con il nome effettivo della tabella Iceberg.

- Per ottenere i dettagli dell'ultima esecuzione di compattazione per una tabella

```
aws get-table-optimizer \
  --catalog-id 123456789012 \
  --database-name iceberg_db \
  --table-name iceberg_table \
  --type compaction
```

- Utilizza l'esempio seguente per recuperare la cronologia di un ottimizzatore per una tabella specifica.

```
aws list-table-optimizer-runs \
  --catalog-id 123456789012 \
  --database-name iceberg_db \
  --table-name iceberg_table \
  --type compaction
```

- L'esempio seguente mostra come recuperare l'esecuzione di compattazione e i dettagli di configurazione per più ottimizzatori. Puoi specificare un massimo di 20 ottimizzatori.

```
aws glue batch-get-table-optimizer \
--entries '[{"catalogId":"123456789012", "databaseName":"iceberg_db",
 "tableName":"iceberg_table", "type":"compaction"}]'
```

AWS API

- Usa l'operazione `GetTableOptimizer` per recuperare i dettagli dell'ultima esecuzione di un ottimizzatore.
- Usa l'operazione `ListTableOptimizerRuns` per recuperare la cronologia di un determinato ottimizzatore su una tabella specifica. È possibile specificare 20 ottimizzatori in una singola chiamata API.
- Usa l'operazione `BatchGetTableOptimizer` per recuperare i dettagli di configurazione per più ottimizzatori nel tuo account. Questa operazione non supporta le chiamate multi-account.

Visualizzazione dei parametri Amazon CloudWatch

Dopo aver eseguito correttamente la compattazione, il servizio crea parametri Amazon CloudWatch sulle prestazioni del processo di compattazione. Puoi andare su CloudWatch Metriche e scegliere Metriche, Tutte le metriche. Puoi filtrare i parametri in base allo spazio dei nomi specifico (ad esempio AWS Glue), al nome della tabella o al nome del database.

Per ulteriori informazioni, consulta Visualizzazione di parametri disponibili nella Guida per l'utente di Amazon CloudWatch.

- Numero di byte compattati
- Numero di file compattati
- Numero di DPU allocato al lavoro
- Durata del processo (ore)

Eliminazione di un ottimizzatore

È possibile eliminare un ottimizzatore e i metadati associati per la tabella utilizzando AWS CLI o un'operazione API AWS.

Esegui il comando AWS CLI seguente per eliminare la cronologia di compattazione per una tabella.

```
aws glue delete-table-optimizer \
  --catalog-id 123456789012 \
  --database-name iceberg_db \
  --table-name iceberg_table \
  --type compaction
```

Usa l'operazione `DeleteTableOptimizer` per eliminare un ottimizzatore per una tabella.

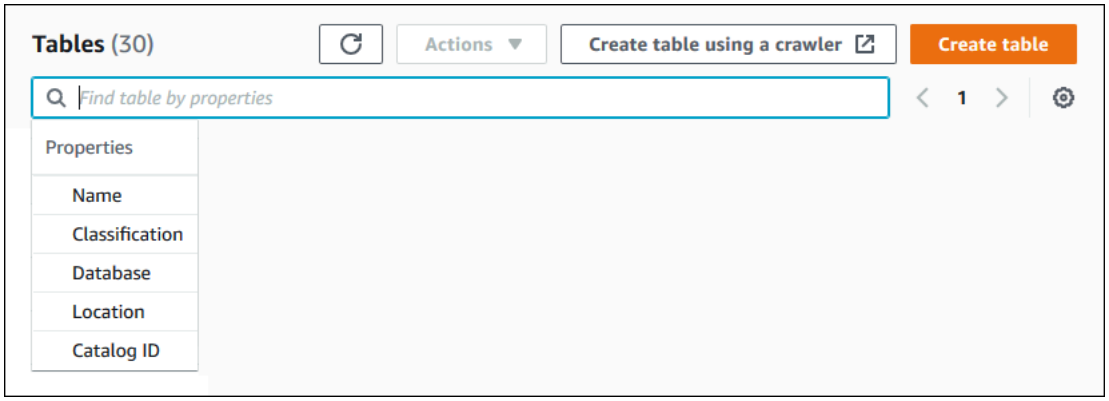
Ricerca di tabelle

Puoi usare la AWS Lake Formation console per cercare le tabelle del Data Catalog per nome, posizione, database contenente e altro. I risultati della ricerca mostrano solo le tabelle per le quali disponi delle autorizzazioni Lake Formation.

Per cercare tabelle (console)

1. Accedi AWS Management Console e apri la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.
2. Nel pannello di navigazione, seleziona Tabelle.
3. Posiziona il cursore nel campo di ricerca nella parte superiore della pagina. Il campo ha il testo segnaposto Trova tabella per proprietà.

   Viene visualizzato il menu Proprietà, che mostra le varie proprietà della tabella in base alle quali effettuare la ricerca.



4. Esegui una di queste operazioni:

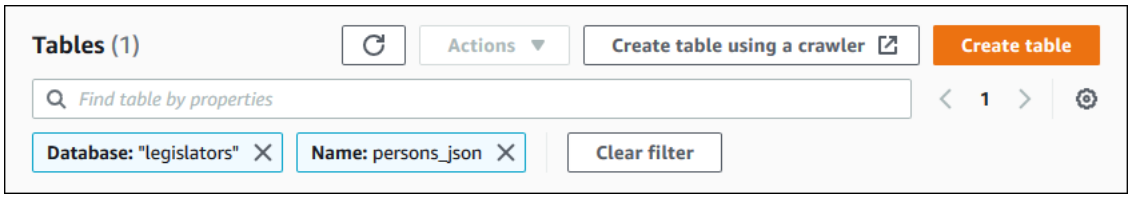
- Ricerca per database contenente.

  1. Scegli Database dal menu Proprietà, quindi scegli un database dal menu Database visualizzato oppure digita il nome del database e premi Invio.

     Vengono elencate le tabelle per le quali disponi delle autorizzazioni nel database.

  2. (Facoltativo) Per restringere l'elenco a una singola tabella del database, posizionate nuovamente il cursore nel campo di ricerca, scegliete Nome dal menu Proprietà e scegliete il nome di una tabella dal menu Tabelle visualizzato oppure digitate un nome di tabella e premete Invio.

     Viene elencata la singola tabella e sia il nome del database che il nome della tabella vengono visualizzati come riquadri sotto il campo di ricerca.

     Per regolare il filtro, chiudi uno dei riquadri o scegli Cancella filtro.

- Cerca per altre proprietà.

  1. Scegliete una proprietà di ricerca dal menu Proprietà.

     Per eseguire la ricerca in base all'ID AWS dell'account, scegliete ID catalogo dal menu Proprietà, immettete un ID AWS account valido (ad esempio, 111122223333) e premete Invio.

Per effettuare una ricerca per località, scegli Posizione dal menu Proprietà e seleziona una località dal menu Posizioni visualizzato. Vengono restituite tutte le tabelle nella posizione principale della posizione selezionata (ad esempio, Amazon S3).

Condivisione di tabelle e database del Data Catalog tra account AWS

Puoi condividere le risorse del Data Catalog (database e tabelle) con AWS account esterni concedendo le autorizzazioni di Lake Formation sulle risorse agli account esterni. Gli utenti possono quindi eseguire query e lavori che uniscono e interrogano tabelle su più account. Con alcune restrizioni, quando condividi una risorsa del Catalogo dati con un altro account, i responsabili di quell'account possono utilizzare quella risorsa come se la risorsa fosse nel loro Catalogo dati.

Non condividi le risorse con responsabili specifici negli AWS account esterni, ma con un account o un'organizzazione. AWS Quando condividi una risorsa con un'AWSorganizzazione, la condividi con tutti gli account a tutti i livelli dell'organizzazione. L'amministratore del data lake di ogni account esterno deve quindi concedere le autorizzazioni sulle risorse condivise ai responsabili del proprio account.

Per ulteriori informazioni, consultare Condivisione dei dati tra account in Lake Formation e Concessione e revoca delle autorizzazioni per le risorse del Data Catalog.

Consulta anche:

- Accesso e visualizzazione di tabelle e database condivisi del Data Catalog
- Prerequisiti

Utilizzo delle visualizzazioni

Questa caratteristica è in versione di anteprima ed è soggetta a modifica. Per ulteriori informazioni, consulta la sezione Beta e anteprime nel documento Termini del servizio AWS.

InAWS Glue Data Catalog, una vista è una tabella virtuale in cui i contenuti sono definiti da una query che fa riferimento a una o più tabelle. Puoi creare una vista che faccia riferimento a un massimo di 10 tabelle utilizzando editor SQL per Amazon Athena, Amazon Redshift o Amazon EMR. Le tabelle di

riferimento sottostanti per una vista possono appartenere allo stesso database o a database diversi all'interno dello stesso. Account AWS

SQL è un linguaggio di programmazione utilizzato per l'interrogazione delle tabelle e ogni motore AWS analitico utilizza la propria variante di SQL, o dialetto SQL. Il Data Catalog supporta la creazione di viste utilizzando dialetti SQL diversi, purché ogni dialetto faccia riferimento allo stesso set di tabelle, colonne e tipi di dati. Definendo uno schema di visualizzazione e un oggetto di metadati comuni che è possibile interrogare da più motori, le viste del catalogo dati consentono di utilizzare viste uniformi in tutto il data lake.

Quando gestisci le viste nel Data Catalog, puoi utilizzarle AWS Lake Formation per concedere autorizzazioni granulari tramite il metodo delle risorse denominate o utilizzando i tag LF e condividerle tra Account AWS organizzazioni e unità organizzative. AWS Puoi anche condividere le viste del Data Catalog tra di loro. Regioni AWS Ciò consente agli utenti di fornire l'accesso ai dati Regioni AWS senza duplicare la fonte dei dati.

Per ulteriori informazioni sulla condivisione dei dati tra account e sull'accesso ai dati tra diverse regioni, consulta:

- Condivisione dei dati tra account in Lake Formation
- Accesso alle tabelle in tutte le regioni

Puoi utilizzare le viste del catalogo dati per:

- Creare e gestire le autorizzazioni su uno schema a visualizzazione singola. Questo ti aiuta a evitare il rischio di autorizzazioni incoerenti su viste duplicate create in più motori.
- Concedi le autorizzazioni agli utenti per una vista che fa riferimento a più tabelle senza concedere le autorizzazioni direttamente sulle tabelle di riferimento sottostanti.

Per le limitazioni, vedere Data Catalog visualizza, considerazioni e limitazioni

Argomenti

- Prerequisiti per la creazione di viste
- Creazione di visualizzazioni
- Concessione delle autorizzazioni per le visualizzazioni del Data Catalog

Prerequisiti per la creazione di viste

- Per creare viste in Data Catalog, devi registrare le posizioni dei dati Amazon S3 sottostanti delle tabelle di riferimento con Lake Formation.

  Per i dettagli sulla registrazione dei dati con Lake Formation, vedereAggiungere una posizione Amazon S3 al tuo data lake.
- Il view definer deve essere un ruolo IAM. Le altre identità IAM non possono creare viste del Data Catalog.
- Il ruolo IAM che definisce la vista deve disporre delle seguenti autorizzazioni:
  - `SELECT`Autorizzazione completa di Lake Formation con `Grantable` opzione su tutte le tabelle di riferimento.
  - Una politica di fiducia affinché Lake Formation e AWS Glue i servizi assumano il ruolo.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "DataCatalogViewDefinerAssumeRole1",
            "Effect": "Allow",
            "Principal": {
               "Service": [
                    "glue.amazonaws.com",
                    "lakeformation.amazonaws.com"
                 ]
            },
            "Action": "sts:AssumeRole"
        }
    ]
}
```

  - Lo scopo: PassRole autorizzazione per AWS Glue e Lake Formation.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "DataCatalogViewDefinerPassRole1",
            "Action": [
```

```
                "iam:PassRole"
            ],
            "Effect": "Allow",
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "iam:PassedToService": [
                        "glue.amazonaws.com",
                        "lakeformation.amazonaws.com"
                      ]
                }
            }
        }
    ]
}
```

- AWS Gluee autorizzazioni Lake Formation.

```
{
    "Version": "2012-10-17",
                "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "Glue:GetDatabase",
                "Glue:GetDatabases",
                "Glue:CreateTable",
                "Glue:GetTable",
                "Glue:UpdateTable",
                "Glue:DeleteTable",
                "Glue:GetTables",
                "Glue:SearchTables",
                "Glue:BatchGetPartition",
                "Glue:GetPartitions",
                "Glue:GetPartition",
                "Glue:GetTableVersion",
                "Glue:GetTableVersions",
                "lakeFormation:GetDataAccess",
                "lakeFormation:GetTemporaryTableCredentials",
                "lakeFormation:GetTemporaryGlueTableCredentials",
                "lakeFormation:GetTemporaryUserCredentialsWithSAML"
            ],
            "Resource": "*"
        }
```

```
    ]
}
```

- Non è possibile creare viste se il database in cui viene creata la vista dispone `Super` o ha l'`ALL`autorizzazione concessa al `IAMAllowedPrincipals` gruppo. Per revocare l'`Super`autorizzazione di un `IAMAllowedPrincipals` gruppo su un database, consulta. Passaggio 4: Passa i tuoi archivi dati al modello di autorizzazioni Lake Formation

  Se le impostazioni del data lake esistenti non ti consentono di impostare il valore `CreateTableDefaultPermissions` vuoto per il `IAMAllowedPrincipals` gruppo, puoi creare un nuovo database e codificare l'impostazione del data lake utilizzando la seguente struttura.

```
{
    "DataLakeSettings": {
        "DataLakeAdmins": [
            {
                "DataLakePrincipalIdentifier":
 "arn:aws:iam::<AccountId>:user/<Username>"
            }
        ],
        CreateTableDefaultPermissions": [
            {
                "Principal": {
                    "DataLakePrincipalIdentifier": "IAM_ALLOWED_PRINCIPALS"
                },
                "Permissions": []
            }
        ]
 }
```

Creazione di visualizzazioni

Puoi utilizzare gli editor SQL per Athena, Amazon Redshift o Amazon EMR per creare viste in. AWS Glue Data Catalog

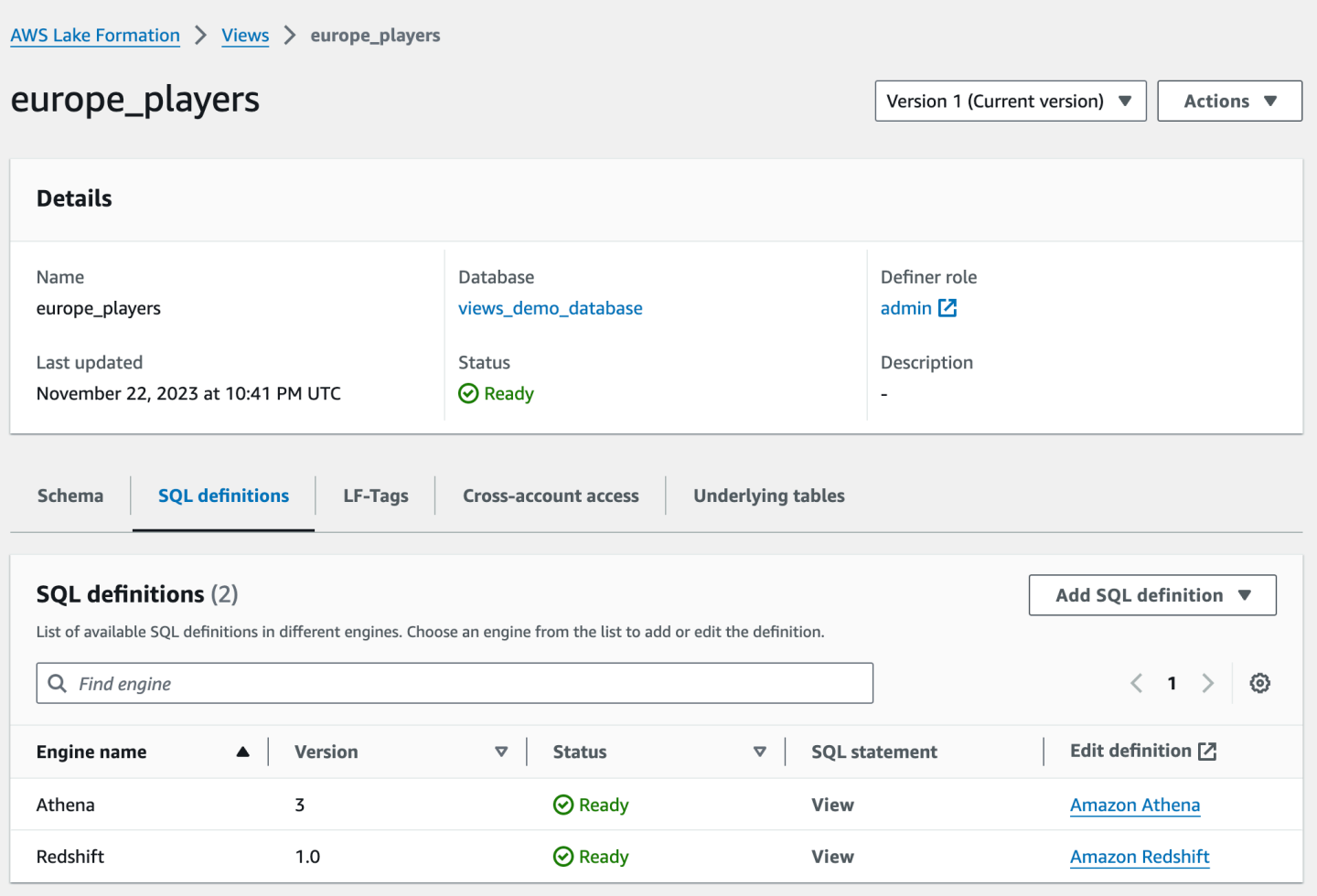
Per ulteriori informazioni sulla sintassi per la creazione e la gestione delle viste del catalogo dati, consulta:

- Utilizzo delle AWS Glue Data Catalog visualizzazioni nella Guida per l'utente di Amazon Athena.
- Creazione di viste AWS Glue Data Catalog nella Amazon Redshift Database Developer Guide.

- Utilizzo delle AWS Glue Data Catalog visualizzazioni nella Amazon EMR Management Guide.

Dopo aver creato una vista Data Catalog, i dettagli della vista nella console Lake Formation.

1. Scegli Views in Data Catalog nella console Lake Formation.
2. Nella pagina delle visualizzazioni viene visualizzato un elenco delle viste disponibili.
3. Scegliete una vista dall'elenco e la pagina dei dettagli mostra gli attributi della vista.



Schema

Scegliete una `Column` riga e selezionate Modifica tag LF per aggiornare i valori dei tag o assegnare nuovi tag LF.

Definizioni SQL

È possibile visualizzare un elenco di definizioni SQL disponibili. Seleziona Aggiungi definizione SQL e scegli un motore di query per aggiungere una definizione SQL. Scegli un motore di query (Athena o Amazon Redshift) `Edit definition` sotto la colonna per aggiornare le definizioni SQL.

Tag LF

Scegliete Modifica tag LF per modificare i valori di un tag o assegnare nuovi tag. È possibile utilizzare i tag LF per concedere autorizzazioni sulle viste.

Accesso multi-account

Puoi visualizzare un elenco di Account AWS organizzazioni e unità organizzative (OU) con cui hai condiviso la visualizzazione del catalogo dati.

Tabelle sottostanti

Le tabelle sottostanti a cui si fa riferimento nella definizione SQL utilizzata per creare la vista sono mostrate in questa scheda.

Concessione delle autorizzazioni per le visualizzazioni del Data Catalog

Dopo aver creato le viste, puoi concedere le autorizzazioni del data lake sulle viste ai responsabili di tutte Account AWS le organizzazioni e le unità organizzative. Per ulteriori informazioni sulla concessione delle autorizzazioni, consulta. Concessione delle autorizzazioni sulle viste utilizzando il metodo di risorsa denominato

Importazione di dati utilizzando i flussi di lavoro in Lake Formation

ConAWS Lake Formation, puoi importare i tuoi dati utilizzando i flussi di lavoro. Un flusso di lavoro definisce l'origine dei dati e la pianificazione per importare i dati nel data lake. È un contenitore per AWS Glue crawler, job e trigger che viene utilizzato per orchestrare i processi di caricamento e aggiornamento del data lake.

Argomenti

- Progetti e flussi di lavoro in Lake Formation
- Creazione di un flusso di lavoro
- Esecuzioni di un flusso di lavoro

Progetti e flussi di lavoro in Lake Formation

Un flusso di lavoro rende possibile incapsulare una complessa attività di estrazione, trasformazione e caricamento (ETL). I flussi di lavoro generano AWS Glue crawler, job e trigger per orchestrare il caricamento e l'aggiornamento dei dati. Lake Formation esegue e traccia un flusso di lavoro come un'unica entità. È possibile configurare un flusso di lavoro per l'esecuzione su richiesta o in base a una pianificazione.

I flussi di lavoro che crei in Lake Formation sono visibili nella AWS Glue console come grafo aciclico orientato (DAG). Ogni nodo DAG è un job, un crawler o un trigger. Per monitorare i progressi e risolvere i problemi, puoi tenere traccia dello stato di ogni nodo del flusso di lavoro.

Quando un flusso di lavoro Lake Formation è completato, all'utente che ha eseguito il flusso di lavoro viene concessa l'`SELECT`autorizzazione Lake Formation sulle tabelle del catalogo dati create dal flusso di lavoro.

Puoi anche creare flussi di lavoro in. AWS Glue Tuttavia, poiché Lake Formation consente di creare un flusso di lavoro da un progetto, la creazione di flussi di lavoro è molto più semplice e automatizzata in Lake Formation. Lake Formation fornisce i seguenti tipi di progetti:

- Istantanea del database: carica o ricarica i dati da tutte le tabelle nel data lake da un'origine JDBC. È possibile escludere alcuni dati dall'origine in base a uno schema di esclusione.
- Database incrementale: carica solo i nuovi dati nel data lake da un'origine JDBC, in base ai segnalibri impostati in precedenza. Si specificano le singole tabelle da includere nel database di origine JDBC. Per ogni tabella, scegli le colonne dei segnalibri e l'ordinamento dei segnalibri per tenere traccia dei dati che sono stati caricati in precedenza. La prima volta che si esegue un blueprint di database incrementale su un set di tabelle, il flusso di lavoro carica tutti i dati dalle tabelle e imposta i segnalibri per la successiva esecuzione del blueprint incrementale del database. È quindi possibile utilizzare un blueprint di database incrementale anziché il blueprint dell'istantanea del database per caricare tutti i dati, a condizione che si specifichi ogni tabella nell'origine dati come parametro.
- File di registro: carica in blocco i dati dalle fonti dei file di registroAWS CloudTrail, inclusi i log di Elastic Load Balancing e i log di Application Load Balancer.

Utilizza la tabella seguente per decidere se utilizzare uno snapshot del database o un blueprint incrementale del database.

Usa l'istantanea del database quando...	Usa il database incrementale quando...
• L'evoluzione dello schema è flessibile. (Le colonne vengono rinominate, le colonne precedenti vengono eliminate e le nuove colonne vengono aggiunte al loro posto.) • È necessaria la coerenza tra l'origine e la destinazione.	• L'evoluzione dello schema è incrementale. (C'è solo l'aggiunta successiva di colonne.) • Vengono aggiunte solo nuove righe; le righe precedenti non vengono aggiornate.

> **Note**
>
> Gli utenti non possono modificare i progetti e i flussi di lavoro creati da Lake Formation.

Creazione di un flusso di lavoro

Prima di iniziare, assicurati di aver concesso le autorizzazioni e le autorizzazioni di localizzazione dei dati richieste per il ruolo. `LakeFormationWorkflowRole` In questo modo il flusso di lavoro può creare tabelle di metadati nel catalogo dati e scrivere dati nelle sedi di destinazione in Amazon S3. Per ulteriori informazioni, consultare (Facoltativo) Crea un ruolo IAM per i flussi di lavoro e Panoramica delle autorizzazioni di Lake Formation .

Per creare un flusso di lavoro da uno schema

1. Apri la AWS Lake Formation console all'indirizzo https://console.aws.amazon.com/lakeformation/. Accedi come amministratore del data lake o come utente con autorizzazioni di data engineer. Per ulteriori informazioni, consulta Riferimento ai personaggi di Lake Formation e alle autorizzazioni IAM.
2. Nel riquadro di navigazione, seleziona I piani, quindi seleziona Usa il piano.
3. Nella pagina Usa un blueprint, scegli un riquadro per selezionare il tipo di blueprint.
4. In Importa origine, specifica l'origine dei dati.

   Se state importando da un'origine JDBC, specificate quanto segue:

- Connessione al database: scegliere una connessione dall'elenco. Crea connessioni aggiuntive utilizzando la AWS Glue console. Il nome utente e la password JDBC nella connessione determinano gli oggetti del database a cui ha accesso il flusso di lavoro.
- Percorso dati di origine: immettere <database><schema>//<table>o<database>/<table>, a seconda del prodotto del database. Oracle Database e MySQL non supportano lo schema nel percorso. È possibile sostituire il carattere di percentuale (%) per *<schema>* o *<table>*. Ad esempio, per un database Oracle con un identificatore di sistema (SID) pari a `orcl`, inserisci `orcl/%` per importare tutte le tabelle a cui ha accesso l'utente indicato nella connessione.

  > **Important**
  >
  > Questo campo distingue tra lettere maiuscole e minuscole Il flusso di lavoro avrà esito negativo in caso di mancata corrispondenza tra maiuscole e minuscole per uno qualsiasi dei componenti.

  Se si specifica un database MySQL, AWS Glue ETL utilizza il driver JDBC Mysql5 per impostazione predefinita, quindi MySQL8 non è supportato in modo nativo. È possibile modificare lo script di lavoro ETL per utilizzare un `customJdbcDriverS3Path` parametro come descritto in JDBC ConnectionType Values nella AWS GlueDeveloper Guide per utilizzare un driver JDBC diverso che supporti MySQL8.

Se stai importando da un file di registro, assicurati che il ruolo specificato per il flusso di lavoro (il «ruolo del flusso di lavoro») disponga delle autorizzazioni IAM necessarie per accedere all'origine dati. Ad esempio, per importare AWS CloudTrail i log, l'utente deve disporre delle autorizzazioni `cloudtrail:DescribeTrails` e delle `cloudtrail:LookupEvents` autorizzazioni per visualizzare l'elenco dei CloudTrail log durante la creazione del flusso di lavoro e il ruolo del flusso di lavoro deve disporre delle autorizzazioni sulla posizione in Amazon S3. CloudTrail

5. Completa una delle seguenti operazioni:

   - Per il tipo di blueprint Database snapshot, identifica facoltativamente un sottoinsieme di dati da importare specificando uno o più modelli di esclusione. Questi modelli di esclusione sono modelli in stile Unix. `glob` Vengono archiviati come proprietà delle tabelle create dal flusso di lavoro.

Per i dettagli sui modelli di esclusione disponibili, consulta Includi ed escludi modelli nella Guida per gli AWS Glue sviluppatori.

- Per il tipo di blueprint del database incrementale, specificare i seguenti campi. Aggiungere una riga per ogni tabella da importare.

  Nome tabella

  Tabella da importare. Deve essere tutto in minuscolo.

  Tasti dei segnalibri

  Elenco delimitato da virgole di nomi di colonne che definiscono le chiavi dei segnalibri. Se è vuota, la chiave primaria viene utilizzata per determinare nuovi dati. Le maiuscole e minuscole per ogni colonna devono corrispondere a quelle definite nell'origine dati.

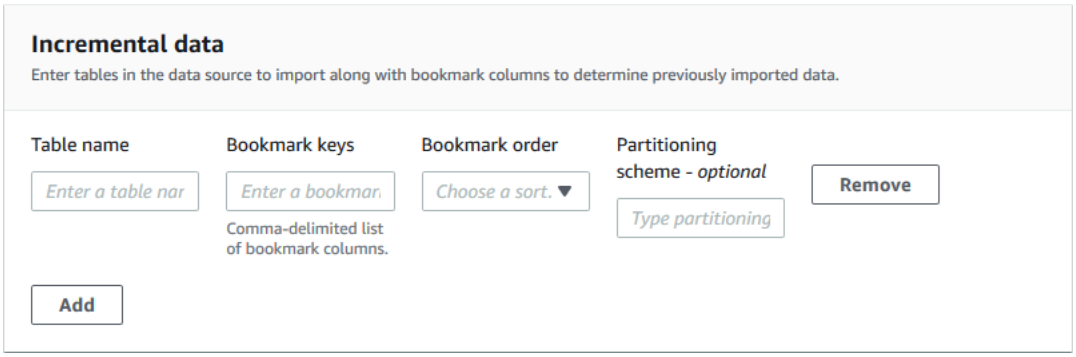
  > **Note**
  >
  > La chiave primaria si qualifica come chiave dei segnalibri predefinita solo se aumenta o diminuisce in sequenza (senza spazi vuoti). Se si desidera utilizzare la chiave primaria come chiave del segnalibro e presenta spazi vuoti, è necessario denominare la colonna chiave primaria come chiave dei segnalibri.

  Ordine dei segnalibri

  Quando scegli Crescente, le righe con valori maggiori dei valori inseriti nei segnalibri vengono identificate come nuove righe. Quando scegli Decrescente, le righe con valori inferiori ai valori dei segnalibri vengono identificate come nuove righe.

  Schema di partizionamento

  (Facoltativo) Elenco delle colonne chiave di partizionamento, delimitate da barre (/). Esempio: `year/month/day`.



Per ulteriori informazioni, consulta Monitoraggio dei dati elaborati tramite segnalibri di Job, consulta la Guida per gli AWS Glue sviluppatori.

6. In Destinazione di importazione, specifica il database di destinazione, la posizione Amazon S3 di destinazione e il formato dei dati.

   Assicurati che il ruolo del flusso di lavoro disponga delle autorizzazioni Lake Formation richieste sul database e sulla posizione di destinazione di Amazon S3.

   > **Note**
   >
   > Attualmente, i blueprint non supportano la crittografia dei dati di destinazione.

7. Scegliete una frequenza di importazione.

   È possibile specificare un'`cron`espressione con l'opzione Personalizzata.

8. In Opzioni di importazione:

   a. Immetti il nome di un flusso di lavoro.

   b. Per il ruolo, scegli il ruolo `LakeFormationWorkflowRole` in cui hai creato(Facoltativo) Crea un ruolo IAM per i flussi di lavoro.

   c. Specifica facoltativamente un prefisso di tabella. Il prefisso viene anteposto ai nomi delle tabelle del catalogo dati create dal flusso di lavoro.

9. Scegli Crea e attendi che la console segnali che il flusso di lavoro è stato creato correttamente.

   > **Tip**
   >
   > È stato visualizzato il messaggio di errore seguente?

```
User: arn:aws:iam::<account-id>:user/<username> is not authorized
to perform: iam:PassRole on resource:arn:aws:iam::<account-
id>:role/<rolename>...
```

In tal caso, verifica di aver sostituito <account-id>con un numero di AWS conto valido in tutte le politiche.

Consulta anche:

- Progetti e flussi di lavoro in Lake Formation

Esecuzioni di un flusso di lavoro

Puoi eseguire un flusso di lavoro utilizzando la console Lake Formation, la AWS Glue console o l'interfaccia a riga di AWS Glue comando (AWS CLI) o l'API.

Per eseguire un flusso di lavoro (console Lake Formation)

1. Apri la AWS Lake Formation console all'indirizzo https://console.aws.amazon.com/lakeformation/. Accedi come amministratore del data lake o come utente con autorizzazioni di data engineer. Per ulteriori informazioni, consulta Riferimento ai personaggi di Lake Formation e alle autorizzazioni IAM.
2. Nel pannello di navigazione scegli Blueprints (Progetti).
3. Nella pagina Blueprints, seleziona il flusso di lavoro. Quindi, nel menu Azioni, scegli Avvia.
4. Durante l'esecuzione del flusso di lavoro, visualizzane l'avanzamento nella colonna Stato dell'ultima esecuzione. Scegliere il pulsante di aggiornamento di tanto in tanto.

   Lo stato va da IN ESECUZIONE, In scoperta, in importazione, a COMPLETATO.

   Quando il flusso di lavoro è completo:

   - Il catalogo dati dispone di nuove tabelle di metadati.
   - I tuoi dati vengono inseriti nel data lake.

   Se il flusso di lavoro fallisce, procedi come segue:

a. Seleziona un flusso di lavoro. Scegli Azioni, quindi scegli Visualizza grafico.

   Il flusso di lavoro si apre nella AWS Glue console.

b. Verifica che il flusso di lavoro sia selezionato e scegli la scheda History (Cronologia).

c. In Cronologia, seleziona la corsa più recente e scegli Visualizza i dettagli della corsa.

d. Seleziona un job o un crawler non riuscito nel grafico dinamico (runtime) ed esamina il messaggio di errore. I nodi non funzionanti sono rossi o gialli.

Consulta anche:

- Progetti e flussi di lavoro in Lake Formation

Gestione delle autorizzazioni di Lake Formation

Lake Formation fornisce controlli di accesso centralizzati per i dati nel tuo data lake. Puoi definire regole basate su policy di sicurezza per i tuoi utenti e le tue applicazioni in base al ruolo in Lake Formation e l'integrazione con AWS Identity and Access Management autentica tali utenti e ruoli. Una volta definite le regole, Lake Formation applica i controlli di accesso a livello di tabella e colonna per gli utenti di Amazon Redshift Spectrum e Amazon Athena.

Argomenti

Concessione delle autorizzazioni per la localizzazione dei dati

Le autorizzazioni di localizzazione dei dati AWS Lake Formation consentono ai responsabili di creare e modificare risorse del catalogo dati che puntano a sedi Amazon S3 registrate designate. Le autorizzazioni di localizzazione dei dati funzionano in aggiunta alle autorizzazioni per i dati di Lake Formation per proteggere le informazioni nel tuo data lake.

Lake Formation non utilizza il servizio AWS Resource Access Manager (AWS RAM) per le concessioni di autorizzazioni alla localizzazione dei dati, quindi non è necessario accettare inviti alla condivisione delle risorse per le autorizzazioni di localizzazione dei dati.

Puoi concedere le autorizzazioni per la localizzazione dei dati utilizzando la console, l'API o AWS Command Line Interface (AWS CLI) di Lake Formation.

> **Note**
>
> Affinché una sovvenzione abbia successo, devi prima registrare la posizione dei dati con Lake Formation.

> **Consulta anche:**
>
> - Underlying data access control

Argomenti

Concessione delle autorizzazioni per la localizzazione dei dati (stesso account)

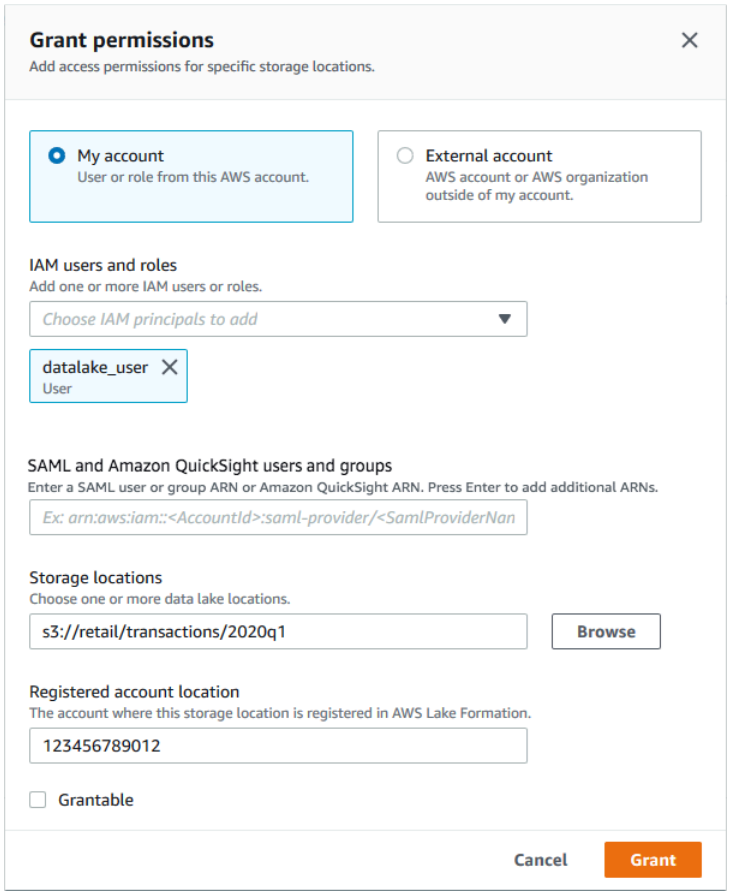
Segui questi passaggi per concedere le autorizzazioni per la localizzazione dei dati ai responsabili del tuo account. AWS Puoi concedere le autorizzazioni utilizzando la console Lake Formation, l'API o AWS Command Line Interface (AWS CLI).

Per concedere le autorizzazioni per la localizzazione dei dati (stesso account, console)

1. Apri la AWS Lake Formation console all'indirizzo https://console.aws.amazon.com/lakeformation/. Accedi come amministratore del data lake o come titolare che dispone delle autorizzazioni concesse per la posizione dei dati desiderata.
2. Nel riquadro di navigazione, scegli Posizioni dei dati.
3. Scegli Concessione.
4. Nella finestra di dialogo Concedi autorizzazioni, assicurati che il riquadro Il mio account sia selezionato. Fornisci quindi le seguenti informazioni:
   - Per gli utenti e i ruoli IAM, scegli uno o più principali.

- Per QuickSight utenti e gruppi SAML e Amazon, inserisci uno o più Amazon Resource Names (ARN) per utenti o gruppi federati tramite SAML o ARNs per utenti o gruppi Amazon. QuickSight

  Immettere un ARN alla volta e premere Invio dopo ogni ARN. Per informazioni su come costruire gli ARN, vedere. Comandi di concessione e AWS CLI revoca di Lake Formation
- Per le posizioni di archiviazione, scegli Browse e scegli una posizione di storage Amazon Simple Storage Service (Amazon S3). La sede deve essere registrata presso Lake Formation. Scegli nuovamente Sfoglia per aggiungere un'altra località. Puoi anche digitare la posizione, ma assicurati di farla precedere da `s3://`.
- Per Ubicazione dell'account registrato, inserisci l'ID AWS dell'account in cui è registrata la sede. L'impostazione predefinita è l'ID del tuo account. In uno scenario con più account, gli amministratori del data lake di un account destinatario possono specificare qui l'account del proprietario quando concedono l'autorizzazione alla localizzazione dei dati ad altri responsabili dell'account del destinatario.
- (Facoltativo) Per consentire ai principali selezionati di concedere le autorizzazioni per la localizzazione dei dati sulla posizione selezionata, seleziona Grantable.



5. Scegli Concessione.

Per concedere le autorizzazioni per la localizzazione dei dati (stesso account,) AWS CLI

- Esegui un `grant-permissions` comando e `DATA_LOCATION_ACCESS` concedi al principale, specificando il percorso Amazon S3 come risorsa.

  Example

  L'esempio seguente concede all'utente le autorizzazioni di localizzazione dei dati. `s3://retail` `datalake_user1`

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/datalake_user1
 --permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"ResourceArn":"arn:aws:s3:::retail"}}'
```

Example

L'esempio seguente concede le autorizzazioni per la localizzazione dei dati a un gruppo. `s3://retail` `ALLIAMPrincipals`

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333:IAMPrincipals --
permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"ResourceArn":"arn:aws:s3:::retail", "CatalogId": "111122223333"}}'
```

Consulta anche:

- Riferimento alle autorizzazioni di Lake Formation

Concessione delle autorizzazioni per la localizzazione dei dati (account esterno)

Segui questi passaggi per concedere le autorizzazioni per la localizzazione dei dati a un AWS account o a un'organizzazione esterni.

Puoi concedere le autorizzazioni utilizzando la console Lake Formation, l'API o AWS Command Line Interface (AWS CLI).

Prima di iniziare

Assicurati che tutti i prerequisiti di accesso tra account siano soddisfatti. Per ulteriori informazioni, consulta Prerequisiti.

Per concedere le autorizzazioni per la localizzazione dei dati (account esterno, console)

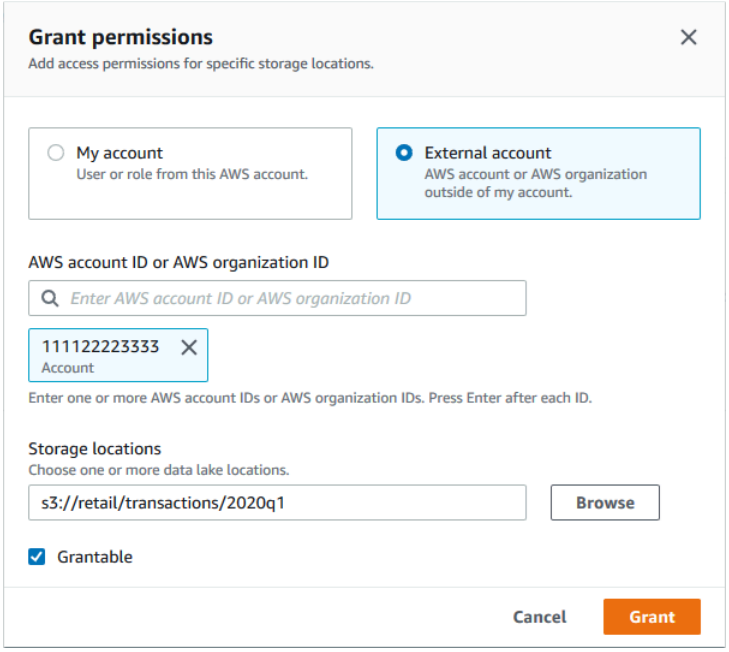
1. Apri la AWS Lake Formation console all'indirizzo https://console.aws.amazon.com/lakeformation/. Accedi come amministratore del data lake.
2. Nel riquadro di navigazione, scegli Posizioni dei dati, quindi scegli Concedi.
3. Nella finestra di dialogo Concedi autorizzazioni, scegli il riquadro Account esterno.
4. Inserisci le informazioni che seguono:

- Per l'ID dell'AWSaccount o AWS l'ID dell'organizzazione, inserisci numeri di AWS conto, ID dell'organizzazione o ID delle unità organizzative validi.

  Premi Invio dopo ogni ID.

  Un ID dell'organizzazione è composto da «o-» seguito da 10-32 lettere o cifre minuscole.

  L'ID di un'unità organizzativa è composto da «ou-» seguito da 4 a 32 lettere o cifre minuscole (l'ID della radice che contiene l'unità organizzativa). Questa stringa è seguita da un secondo «-» (trattino) e da 8 a 32 lettere o cifre minuscole aggiuntive.
- In Luoghi di archiviazione, scegli Browse e scegli una posizione di storage Amazon Simple Storage Service (Amazon S3). La sede deve essere registrata presso Lake Formation.



5. Seleziona Grantable.
6. Scegli Concessione.

Per concedere le autorizzazioni per la localizzazione dei dati (account esterno,) AWS CLI

- Per concedere le autorizzazioni a un AWS account esterno, inserisci un comando simile al seguente.

```
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"}}'
```

Questo comando concede l'opzione di concessione `DATA_LOCATION_ACCESS` all'account 1111-2222-3333 nella posizione `s3://retail/transactions/2020q1` Amazon S3, che è di proprietà dell'account 1234-5678-9012.

Per concedere le autorizzazioni a un'organizzazione, inserisci un comando simile al seguente.

```
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"}}'
```

Questo comando concede un'opzione `DATA_LOCATION_ACCESS` di concessione all'organizzazione `o-abcdefghijkl` nella `s3://retail/transactions/2020q1` posizione Amazon S3, che è di proprietà dell'account 1234-5678-9012.

Per concedere le autorizzazioni a un principale in un AWS account esterno, inserisci un comando simile al seguente.

```
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"}}'
```

Questo comando concede `DATA_LOCATION_ACCESS` a un principale nell'account 1111-2222-3333 nella posizione `s3://retail/transactions/2020q1` Amazon S3, che è di proprietà dell'account 1234-5678-9012.

Example

L'esempio seguente concede le autorizzazioni di localizzazione dei dati per il raggruppamento in un account esterno. `s3://retail` `ALLIAMPrincipals`

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333:IAMPrincipals --
permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"ResourceArn":"arn:aws:s3:::retail", "CatalogId": "123456789012"}}'
```

Consulta anche:

- Riferimento alle autorizzazioni di Lake Formation

Concessione delle autorizzazioni su una posizione di dati condivisa con l'account

Dopo aver condiviso una risorsa Data Catalog con il tuo AWS account, in qualità di amministratore del data lake, puoi concedere le autorizzazioni sulla risorsa ad altri responsabili del tuo account. Se l'`ALTER`autorizzazione viene concessa su una tabella condivisa e la tabella punta a una posizione Amazon S3 registrata, devi anche concedere le autorizzazioni per la posizione dei dati sulla posizione. Allo stesso modo, se l'`ALTER`autorizzazione `CREATE_TABLE` o è concessa a un database condiviso e il database ha una proprietà di location che punta a una posizione registrata, devi anche concedere le autorizzazioni per la localizzazione dei dati sulla posizione.

Per concedere le autorizzazioni di localizzazione dei dati su una posizione condivisa a un responsabile del tuo account, al tuo account deve essere stata concessa l'`DATA_LOCATION_ACCESS`autorizzazione sulla posizione con l'opzione di concessione. Quando poi concedi `DATA_LOCATION_ACCESS` a un altro titolare del tuo account, devi includere l'ID Data Catalog (ID dell'AWSaccount) dell'account proprietario. L'account del proprietario è l'account che ha registrato la sede.

Puoi utilizzare la AWS Lake Formation console, l'API o AWS Command Line Interface () AWS CLI per concedere le autorizzazioni per la localizzazione dei dati.

Per concedere le autorizzazioni su una posizione dati condivisa con il tuo account (console)

- Seguire la procedura riportata in Concessione delle autorizzazioni per la localizzazione dei dati (stesso account).

  Per le posizioni di archiviazione, è necessario digitare le posizioni. Per Ubicazione dell'account registrato, inserisci l'AWSID dell'account del proprietario.

Per concedere autorizzazioni su una posizione dati condivisa con il tuo account () AWS CLI

- Inserisci uno dei seguenti comandi per concedere le autorizzazioni a un utente o a un ruolo.

```
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>"}}'
```

Concessione e revoca delle autorizzazioni per le risorse del Data Catalog

Puoi concedere le autorizzazioni del Data Lake ai responsabili in AWS Lake Formation modo che i responsabili possano creare e gestire le risorse del Data Catalog e possano accedere ai dati sottostanti. È possibile concedere autorizzazioni Data lake su database, tabelle e viste. Quando concedi le autorizzazioni per le tabelle, puoi limitare l'accesso a colonne o righe specifiche della tabella per un controllo degli accessi ancora più preciso.

È possibile concedere autorizzazioni per singole tabelle e viste oppure, con un'unica operazione di concessione, è possibile concedere autorizzazioni su tutte le tabelle e le viste di un database. Se concedi autorizzazioni su tutte le tabelle di un database, concedi implicitamente le autorizzazioni per il `DESCRIBE` database. Il database viene quindi visualizzato nella pagina Database della console e viene restituito dall'`GetDatabases`operazione API.

È possibile concedere le autorizzazioni utilizzando il metodo named resource o il metodo di controllo degli accessi basato su tag Lake Formation (LF-TBAC).

È possibile concedere le autorizzazioni ai responsabili dello stesso account o a organizzazioni esterne. Account AWS Quando concedi ad account o organizzazioni esterni, condividi le risorse che possiedi con tali account o organizzazioni. I responsabili di tali account o organizzazioni possono quindi accedere alle risorse di Data Catalog di tua proprietà e ai dati sottostanti.

> **Note**
>
> Attualmente, il metodo LF-TBAC supporta la concessione di autorizzazioni su più account a dirigenti, Account AWS organizzazioni e unità organizzative (OU) IAM.

Quando concedi autorizzazioni ad account o organizzazioni esterni, devi includere l'opzione di concessione. Solo l'amministratore del data lake dell'account esterno può accedere alle risorse condivise finché l'amministratore non concede le autorizzazioni sulle risorse condivise ad altri responsabili dell'account esterno.

Puoi concedere le autorizzazioni di Data Catalog utilizzando la AWS Lake Formation console, l'API o (). AWS Command Line Interface AWS CLI

> **Note**
>
> Quando elimini una risorsa Data Catalog, tutte le autorizzazioni associate alla risorsa diventano non valide. Ricreando la stessa risorsa con lo stesso nome, non verranno recuperate le autorizzazioni di Lake Formation. Gli utenti dovranno configurare nuovamente le nuove autorizzazioni.

> **Consulta anche:**
>
> - Condivisione di tabelle e database del Data Catalog tra account AWS
> - Controllo dell'accesso ai metadati
> - Riferimento alle autorizzazioni di Lake Formation

Autorizzazioni IAM necessarie per concedere o revocare le autorizzazioni di Lake Formation

Tutti i responsabili, incluso l'amministratore del data lake, necessitano delle seguenti autorizzazioni AWS Identity and Access Management (IAM) per concedere o revocare le autorizzazioni AWS Lake Formation Data Catalog o le autorizzazioni di localizzazione dei dati con l'API Lake Formation o il: AWS CLI

- `lakeformation:GrantPermissions`
- `lakeformation:BatchGrantPermissions`
- `lakeformation:RevokePermissions`
- `lakeformation:BatchRevokePermissions`
- `glue:GetTable`o `glue:GetDatabase` per una tabella o un database a cui stai concedendo le autorizzazioni utilizzando il metodo della risorsa denominata.

> **Note**
>
> Gli amministratori di Data Lake dispongono delle autorizzazioni implicite di Lake Formation per concedere e revocare le autorizzazioni di Lake Formation. Ma hanno comunque bisogno delle autorizzazioni IAM sulle operazioni API di concessione e revoca di Lake Formation. I ruoli IAM con policy `AWSLakeFormationDataAdmin` AWS gestita non possono aggiungere nuovi amministratori di data lake perché questa policy contiene un rifiuto esplicito per il funzionamento dell'API Lake Formation,. `PutDataLakeSetting`

La seguente policy IAM è consigliata ai responsabili che non sono amministratori di data lake e che desiderano concedere o revocare le autorizzazioni utilizzando la console Lake Formation.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:ListPermissions",
                "lakeformation:GrantPermissions",
                "lakeformation:BatchGrantPermissions",
```

```
                "lakeformation:RevokePermissions",
                "lakeformation:BatchRevokePermissions",
                "glue:GetDatabases",
                "glue:SearchTables",
                "glue:GetTables",
                "glue:GetDatabase",
                "glue:GetTable",
                "iam:ListUsers",
                "iam:ListRoles",
                "sso-directory:DescribeUser",
                "sso-directory:DescribeGroup",
                "sso:DescribeInstance"
            ],
            "Resource": "*"
        }
    ]
}
```

Tutte le autorizzazioni `glue:` e le `iam:` autorizzazioni di questa policy sono disponibili nella policy gestita. AWS `AWSGlueConsoleFullAccess`

Per concedere le autorizzazioni utilizzando il controllo degli accessi basato su tag Lake Formation (LF-TBAC), i principali necessitano di autorizzazioni IAM aggiuntive. Per ulteriori informazioni, consultare Buone pratiche e considerazioni per il controllo degli accessi basato su tag Lake Formation e Riferimento ai personaggi di Lake Formation e alle autorizzazioni IAM.

Autorizzazioni tra account

Gli utenti che desiderano concedere autorizzazioni Lake Formation su più account utilizzando il metodo della risorsa denominata devono disporre anche delle autorizzazioni nella `AWSLakeFormationCrossAccountManager` AWS politica gestita.

Gli amministratori di Data Lake necessitano delle stesse autorizzazioni per concedere autorizzazioni su più account, oltre all'autorizzazione AWS Resource Access Manager (AWS RAM) per consentire la concessione di autorizzazioni alle organizzazioni. Per ulteriori informazioni, consulta Autorizzazioni di amministratore di Data Lake.

L'utente amministrativo

Un preside con autorizzazioni amministrative, ad esempio con la policy `AdministratorAccess` AWS gestita, dispone delle autorizzazioni per concedere le autorizzazioni di Lake Formation e creare amministratori di data lake. Per negare a un utente o a un ruolo l'accesso alle operazioni di

amministratore di Lake Formation, allega o aggiungi alla sua policy una `Deny` dichiarazione per le operazioni dell'API dell'amministratore.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "lakeformation:GetDataLakeSettings",
                "lakeformation:PutDataLakeSettings"
            ],
            "Effect": "Deny",
            "Resource": [
                "*"
            ]
        }
    ]
}
```

> **Important**
>
> Per impedire agli utenti di aggiungersi come amministratori con uno script di estrazione, trasformazione e caricamento (ETL), assicurati che a tutti gli utenti e i ruoli non amministratori sia negato l'accesso a queste operazioni API. La policy `AWSLakeFormationDataAdmin` AWS gestita contiene una negazione esplicita del funzionamento dell'API Lake Formation, `PutDataLakeSetting` che impedisce agli utenti di aggiungere nuovi amministratori di data lake.

Concessione delle autorizzazioni per il data lake utilizzando il metodo di risorsa denominato

Puoi utilizzare il metodo della risorsa denominata per concedere le autorizzazioni di Lake Formation su database, tabelle e viste specifici del Data Catalog. Puoi concedere le autorizzazioni utilizzando la AWS Lake Formation console, l'API o AWS Command Line Interface ()AWS CLI.

Argomenti

- Concessione delle autorizzazioni al database utilizzando il metodo di risorsa denominato
- Concessione delle autorizzazioni per le tabelle utilizzando il metodo di risorsa denominato

- Concessione delle autorizzazioni sulle viste utilizzando il metodo di risorsa denominato

Concessione delle autorizzazioni al database utilizzando il metodo di risorsa denominato

I passaggi seguenti spiegano come concedere le autorizzazioni al database utilizzando il metodo della risorsa denominata.

Console

Usa la pagina Concedi le autorizzazioni del data lake sulla console Lake Formation. La pagina è suddivisa nelle seguenti sezioni:

- Principi: utenti, ruoli, utenti e gruppi IAM Identity Center, utenti e gruppi SAML, AWS account, organizzazioni o unità organizzative a cui concedere le autorizzazioni.
- Tag LF o risorse del catalogo: database, tabelle, viste o collegamenti alle risorse su cui concedere le autorizzazioni.
- Autorizzazioni: autorizzazioni da concedere a The Lake Formation.

> **Note**
>
> Per concedere le autorizzazioni su un collegamento a una risorsa del database, vedere. Concessione delle autorizzazioni per i collegamenti alle risorse

1. Apri la pagina Concedi le autorizzazioni del data lake.

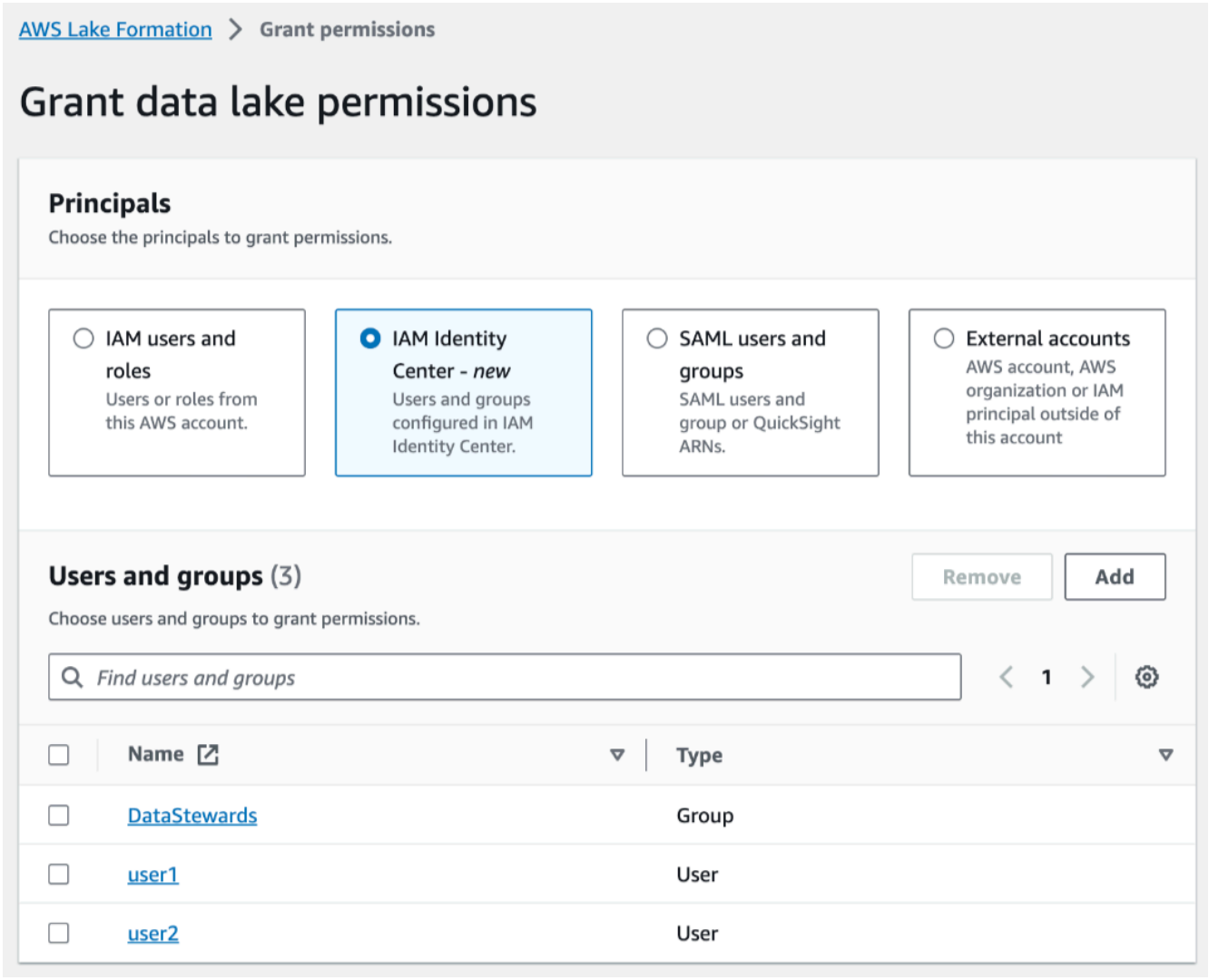
   Apri la AWS Lake Formation console all'indirizzo https://console.aws.amazon.com/lakeformation/ e accedi come amministratore del data lake, creatore del database o utente IAM con autorizzazioni Grantable sul database.

   Esegui una di queste operazioni:

   - Nel riquadro di navigazione, in Autorizzazioni, scegli Autorizzazioni Data Lake. Quindi scegli Concedi.
   - Nel riquadro di navigazione, scegli Database in Data Catalog. Quindi, nella pagina Database, scegli un database e dal menu Azioni, in Autorizzazioni, scegli Concedi.

> **Note**
>
> Puoi concedere le autorizzazioni su un database tramite il relativo link alla risorsa. A tale scopo, nella pagina Database, scegli un collegamento a una risorsa e nel menu Azioni scegli Concedi alla destinazione. Per ulteriori informazioni, consulta Come funzionano i link alle risorse in Lake Formation.

2. Successivamente, nella sezione Principi, scegli un tipo principale e quindi specifica i principali a cui concedere le autorizzazioni.



Utenti e ruoli IAM

Scegli uno o più utenti o ruoli dall'elenco degli utenti e dei ruoli IAM.

IAM Identity Center

Scegli uno o più utenti o gruppi dall'elenco Utenti e gruppi. Seleziona Aggiungi per aggiungere altri utenti o gruppi.

Utenti e gruppi SAML

Per QuickSight utenti e gruppi SAML e Amazon, inserisci uno o più Amazon Resource Names (ARN) per utenti o gruppi federati tramite SAML o ARN per utenti o gruppi Amazon. QuickSight Premi Invio dopo ogni ARN.

Per informazioni su come costruire gli ARN, vedere. Comandi di concessione e AWS CLI revoca di Lake Formation

> **Note**
>
> L'integrazione di Lake Formation con Amazon QuickSight è supportata solo per Amazon QuickSight Enterprise Edition.

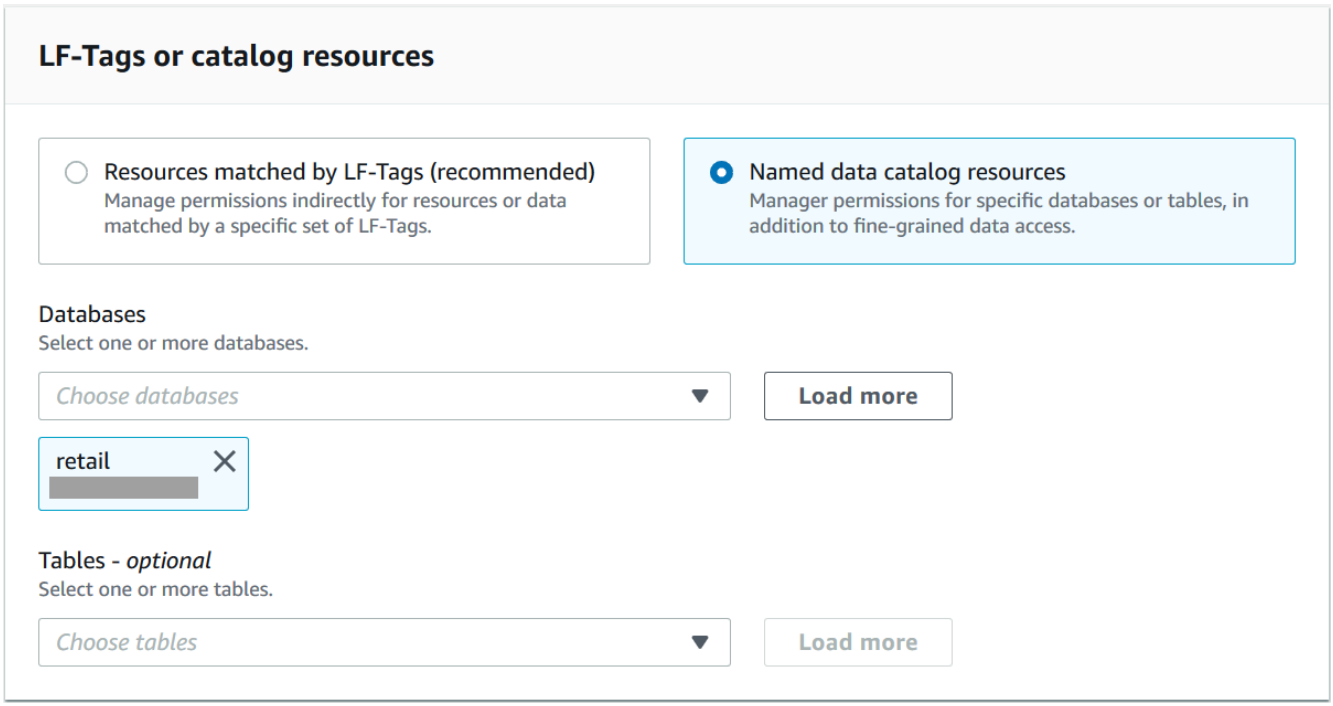
Account esterni

Per Account AWS, AWS organizzazione o Principal IAM inserisci uno o più ID AWS account, ID organizzazione, ID unità organizzative o ARN validi per l'utente o il ruolo IAM. Premi Invio dopo ogni ID.

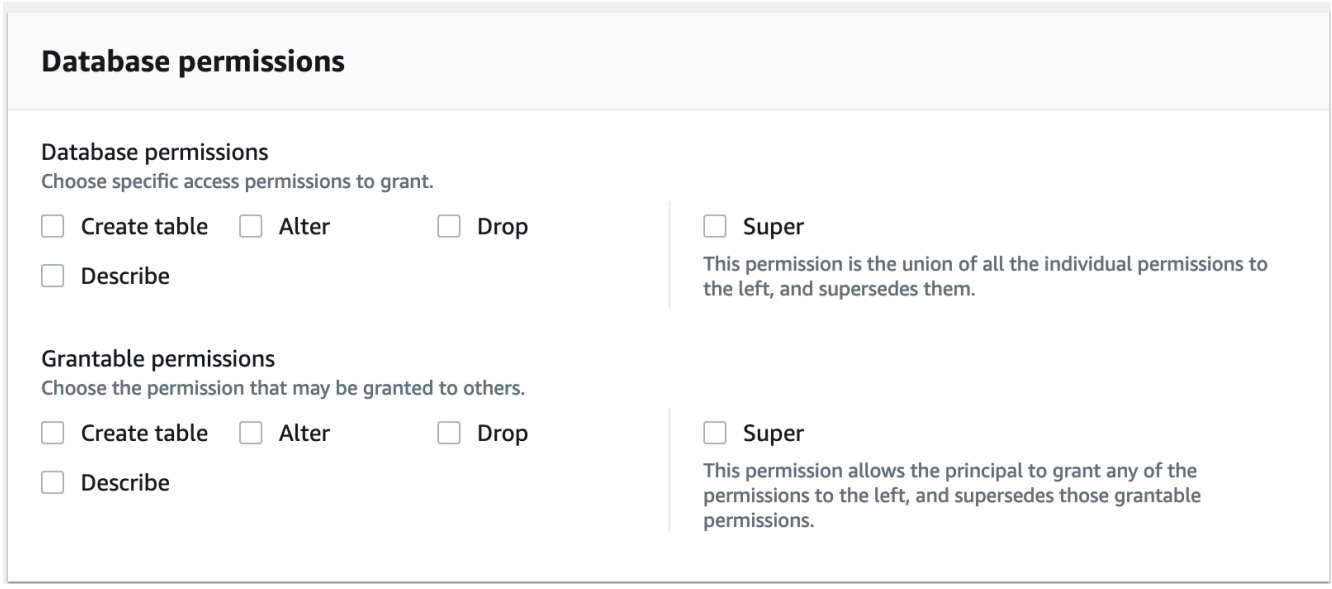
Un ID dell'organizzazione è composto da «o-» seguito da 10-32 lettere o cifre minuscole.

L'ID di un'unità organizzativa inizia con «ou-» seguito da 4—32 lettere o cifre minuscole (l'ID della radice che contiene l'unità organizzativa). Questa stringa è seguita da un secondo trattino «-» e da 8 a 32 lettere o cifre minuscole aggiuntive.

3. Nella sezione LF-Tags o catalog resources, scegliete Named data catalog resources.



4. Scegli uno o più database dall'elenco Database. Puoi anche scegliere una o più tabelle e/o filtri di dati.
5. Nella sezione Autorizzazioni, seleziona autorizzazioni e autorizzazioni concedibili. In Autorizzazioni del database, seleziona una o più autorizzazioni da concedere.



> **Note**
>
> Dopo aver concesso `Create Table` o `Alter` su un database con una proprietà location che punta a una posizione registrata, assicurati di concedere anche ai

responsabili le autorizzazioni per la localizzazione dei dati sulla posizione. Per ulteriori informazioni, consulta Concessione delle autorizzazioni per la localizzazione dei dati.

6. (Facoltativo) In Autorizzazioni concedibili, seleziona le autorizzazioni che il destinatario della sovvenzione può concedere ad altri responsabili del proprio account. AWS Questa opzione non è supportata quando si concedono autorizzazioni a un responsabile IAM da un account esterno.
7. Scegli Concessione.

AWS CLI

È possibile concedere le autorizzazioni al database utilizzando il metodo di risorsa denominato e il AWS Command Line Interface (). AWS CLI

Per concedere le autorizzazioni al database utilizzando AWS CLI

- Eseguite un `grant-permissions` comando e specificate un database o il Data Catalog come risorsa, a seconda dell'autorizzazione concessa.

  Negli esempi seguenti, sostituiscilo <account-id>con un ID AWS account valido.

  Example — Concedi la creazione di un database

  Questo esempio concede `CREATE_DATABASE` all'utente`datalake_user1`. Poiché la risorsa su cui viene concessa questa autorizzazione è il Data Catalog, il comando specifica una `CatalogResource` struttura vuota come parametro. `resource`

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/datalake_user1 --
permissions "CREATE_DATABASE" --resource '{ "Catalog": {}}'
```

  Example — Concedi la creazione di tabelle in un database designato

  L'esempio successivo concede l'`CREATE_TABLE`accesso al database `retail` all'utente`datalake_user1`.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/datalake_user1 --
permissions "CREATE_TABLE" --resource '{ "Database": {"Name":"retail"}}'
```

Example — Concedi a un AWS account esterno con l'opzione Grant

L'esempio successivo concede `CREATE_TABLE` con l'opzione grant sul database `retail` all'account esterno 1111-2222-3333.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333 --permissions "CREATE_TABLE"
 --permissions-with-grant-option "CREATE_TABLE" --resource '{ "Database":
 {"Name":"retail"}}'
```

Example — Concessione a un'organizzazione

L'esempio successivo concede all'organizzazione `ALTER` `o-abcdefghijkl` con l'opzione di concessione sul database`issues`.

```
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 - Concedi a **`ALLIAMPrincipals`** nello stesso account

L'esempio successivo concede l'`CREATE_TABLE`autorizzazione sul database `retail` a tutti i responsabili dello stesso account. Questa opzione consente a tutti i principali dell'account di creare una tabella nel database e creare un collegamento alle risorse della tabella che consente ai motori di query integrati di accedere a database e tabelle condivisi. Questa opzione è particolarmente utile quando un principale riceve una sovvenzione tra account e non dispone dell'autorizzazione per creare collegamenti alle risorse. In questo scenario, l'amministratore del data lake può creare un database segnaposto e concedere l'`CREATE_TABLE`autorizzazione al `ALLIAMPrincipal` gruppo, consentendo a ogni principale IAM dell'account di creare collegamenti alle risorse nel database placeholder.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333:IAMPrincipals
```

```
 --permissions "CREATE_TABLE"  --resource '{ "Database":
 {"Name":"temp","CatalogId":"111122223333"}}'
```

Example - Concedi a **`ALLIAMPrincipals`** in un account esterno

L'esempio successivo concede l'`CREATE_TABLE`accesso al database `retail` a tutti i principali di un account esterno. Questa opzione consente a tutti i principali dell'account di creare una tabella nel database.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333:IAMPrincipals
 --permissions "CREATE_TABLE"  --resource '{ "Database":
 {"Name":"retail","CatalogId":"123456789012"}}'
```

> **Note**
>
> Dopo aver concesso `CREATE_TABLE` o `ALTER` su un database con una proprietà location che punta a una posizione registrata, assicurati di concedere anche ai responsabili le autorizzazioni di localizzazione dei dati sulla posizione. Per ulteriori informazioni, consulta Concessione delle autorizzazioni per la localizzazione dei dati.

> **Consulta anche**
>
> - Riferimento alle autorizzazioni di Lake Formation
> - Concessione delle autorizzazioni su un database o una tabella condivisa con il tuo account
> - Accesso e visualizzazione di tabelle e database condivisi del Data Catalog

Concessione delle autorizzazioni per le tabelle utilizzando il metodo di risorsa denominato

Puoi utilizzare la console Lake Formation o AWS CLI concedere le autorizzazioni di Lake Formation sulle tabelle di Data Catalog. È possibile concedere autorizzazioni su singole tabelle oppure con una singola operazione di concessione, è possibile concedere autorizzazioni su tutte le tabelle di un database.

Se concedi le autorizzazioni per tutte le tabelle di un database, concedi implicitamente le autorizzazioni per il `DESCRIBE` database. Il database viene quindi visualizzato nella pagina Database della console e viene restituito dall'`GetDatabases`operazione API.

Quando scegli `SELECT` come autorizzazione da concedere, hai la possibilità di applicare un filtro di colonna, un filtro di riga o un filtro di cella.

Console

I passaggi seguenti spiegano come concedere le autorizzazioni alle tabelle utilizzando il metodo della risorsa denominata e la pagina Grant data lake permissions sulla console Lake Formation. La pagina è suddivisa nelle seguenti sezioni:

- Principi: utenti, ruoli, AWS account, organizzazioni o unità organizzative a cui concedere le autorizzazioni.
- LF-Tag o risorse di catalogo: database, tabelle o collegamenti a risorse a cui concedere le autorizzazioni.
- Autorizzazioni: autorizzazioni da concedere a The Lake Formation.

> **Note**
>
> Per concedere le autorizzazioni su un collegamento a una risorsa della tabella, vedere. Concessione delle autorizzazioni per i collegamenti alle risorse

1. Apri la pagina Concedi le autorizzazioni del data lake.

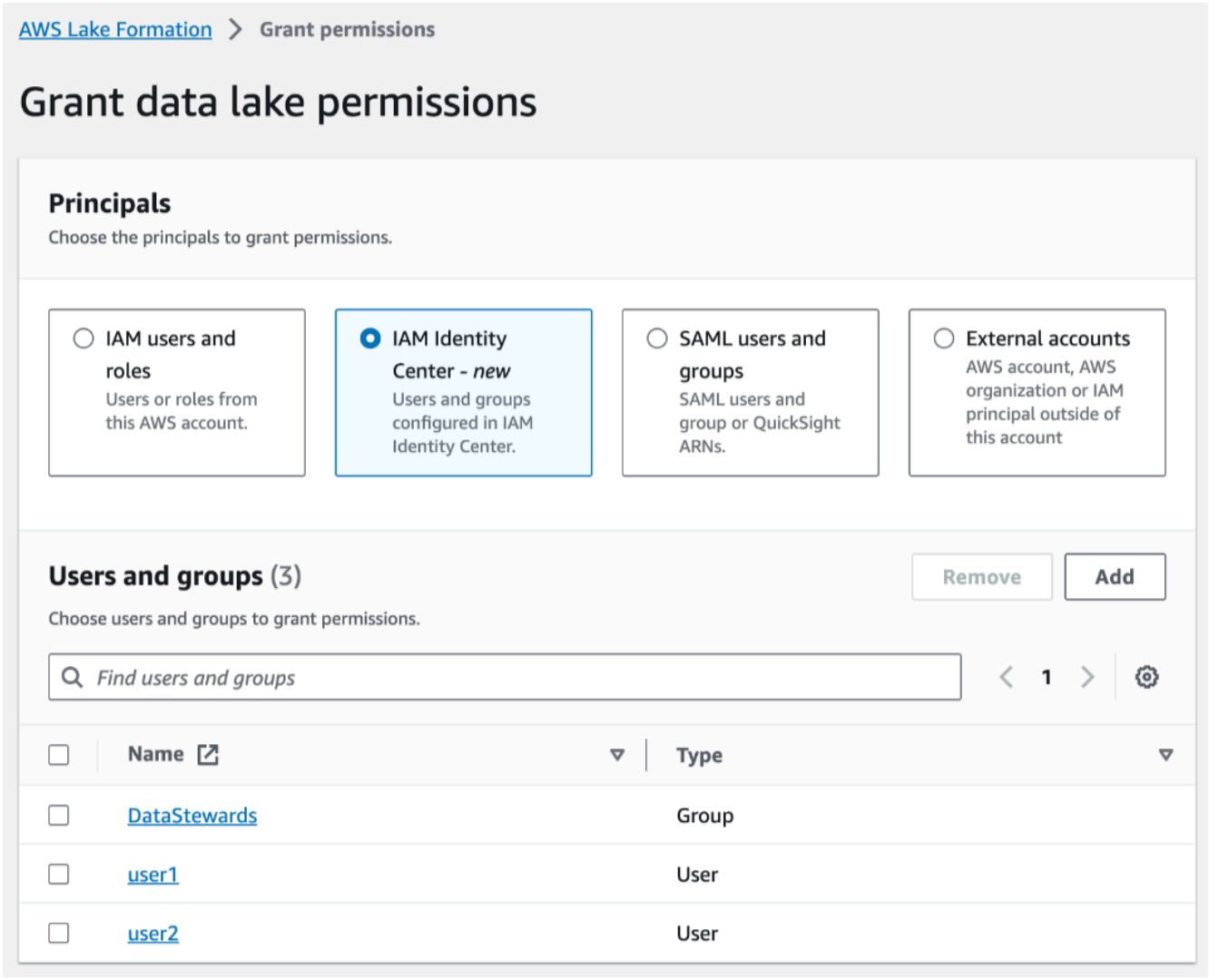
   Apri la AWS Lake Formation console all'indirizzo https://console.aws.amazon.com/lakeformation/ e accedi come amministratore del data lake, creatore della tabella o utente a cui sono state concesse le autorizzazioni sulla tabella con l'opzione grant.

   Esegui una di queste operazioni:

   - Nel riquadro di navigazione, scegli Autorizzazioni Data lake in Autorizzazioni. Quindi scegli Concedi.
   - Nel pannello di navigazione, seleziona Tabelle. Quindi, nella pagina Tabelle, scegli una tabella e nel menu Azioni, in Autorizzazioni, scegli Concedi.

> **Note**
>
> Puoi concedere le autorizzazioni su una tabella tramite il relativo link alla risorsa. A tale scopo, nella pagina Tabelle, scegli un collegamento a una risorsa e nel menu Azioni scegli Concedi all'obiettivo. Per ulteriori informazioni, consulta Come funzionano i link alle risorse in Lake Formation.

2. Successivamente, nella sezione Principi, scegli un tipo principale e specifica i principali a cui concedere le autorizzazioni.



Utenti e ruoli IAM

Scegli uno o più utenti o ruoli dall'elenco degli utenti e dei ruoli IAM.

IAM Identity Center

Scegli uno o più utenti o gruppi dall'elenco Utenti e gruppi.

Utenti e gruppi SAML

Per QuickSight utenti e gruppi SAML e Amazon, inserisci uno o più Amazon Resource Names (ARN) per utenti o gruppi federati tramite SAML o ARN per utenti o gruppi Amazon. QuickSight Premi Invio dopo ogni ARN.

Per informazioni su come costruire gli ARN, vedere. Comandi di concessione e AWS CLI revoca di Lake Formation

> **Note**
>
> L'integrazione di Lake Formation con Amazon QuickSight è supportata solo per Amazon QuickSight Enterprise Edition.

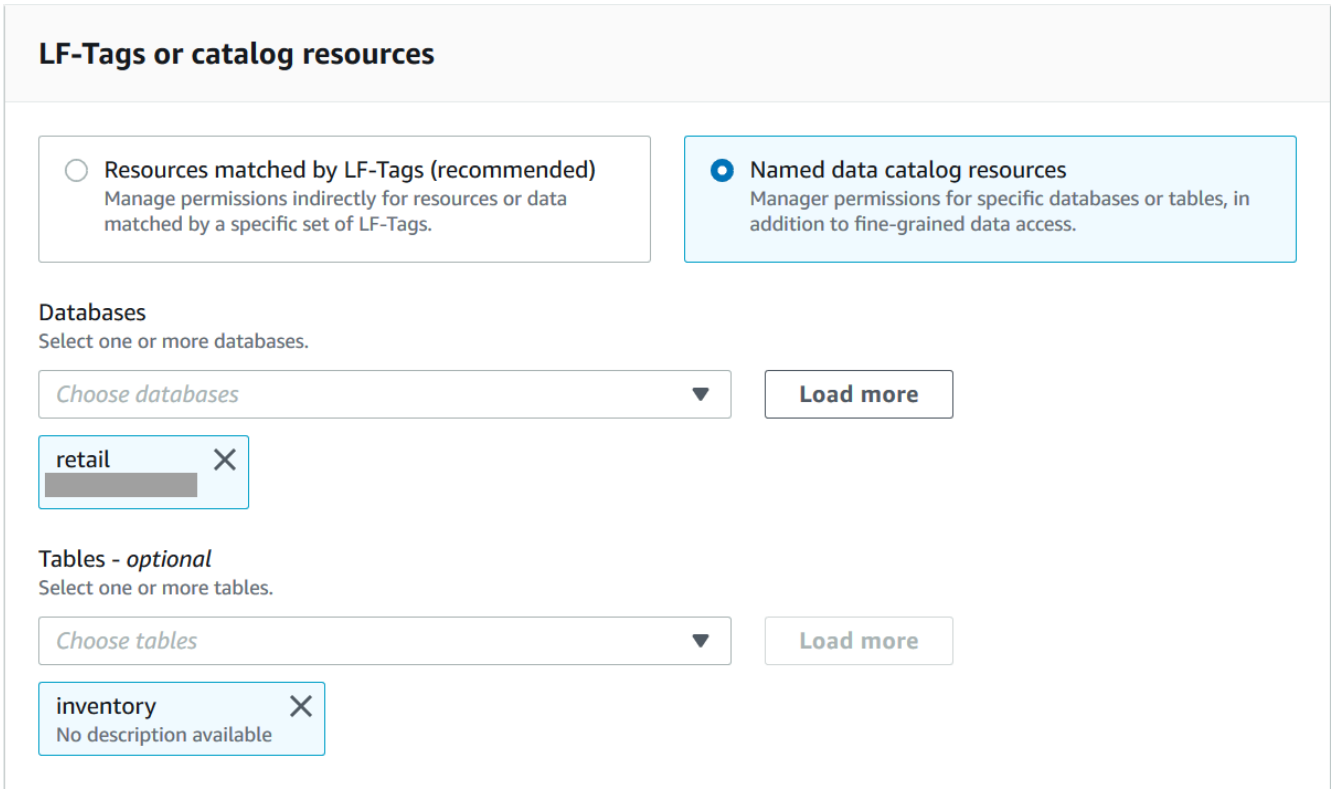
Account esterni

Per Account AWS, AWS organizzazione o Principal IAM inserisci uno o più ID, ID di organizzazione, ID di unità organizzative Account AWS validi o l'ARN per l'utente o il ruolo IAM. Premi Invio dopo ogni ID.

Un ID dell'organizzazione è composto da «o-» seguito da 10-32 lettere o cifre minuscole.

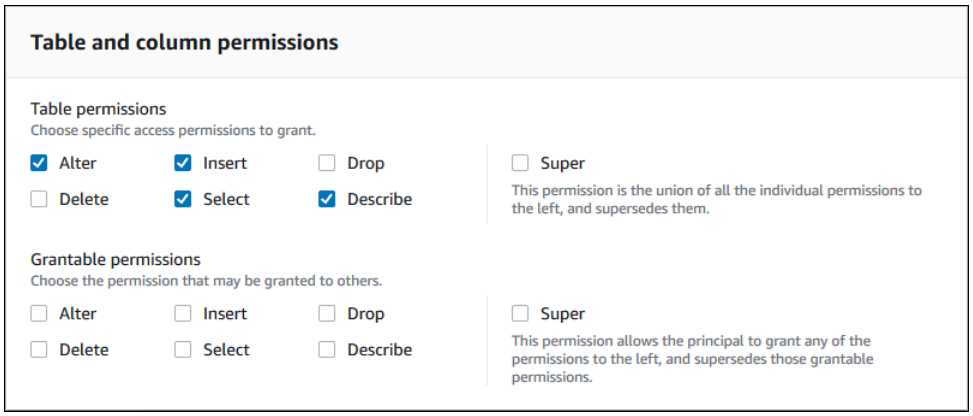
L'ID di un'unità organizzativa inizia con «ou-» seguito da 4—32 lettere o cifre minuscole (l'ID della radice che contiene l'unità organizzativa). Questa stringa è seguita da un secondo carattere «-» e da 8 a 32 lettere o cifre minuscole aggiuntive.

3. Nella sezione LF-tags o risorse del catalogo, scegliete un database. Quindi scegliete una o più tabelle o Tutte le tabelle.

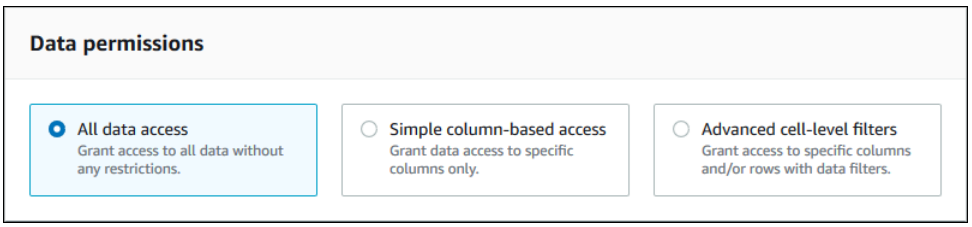


4. Specificate le autorizzazioni senza filtraggio dei dati

   Nella sezione Autorizzazioni, seleziona la tabella autorizzazioni da concedere e, facoltativamente, seleziona le autorizzazioni concedibili.



   Se concedi Select, la sezione Autorizzazioni ai dati viene visualizzata sotto la sezione Autorizzazioni per tabelle e colonne, con l'opzione di accesso a tutti i dati selezionata per impostazione predefinita. Accetta l'impostazione predefinita.



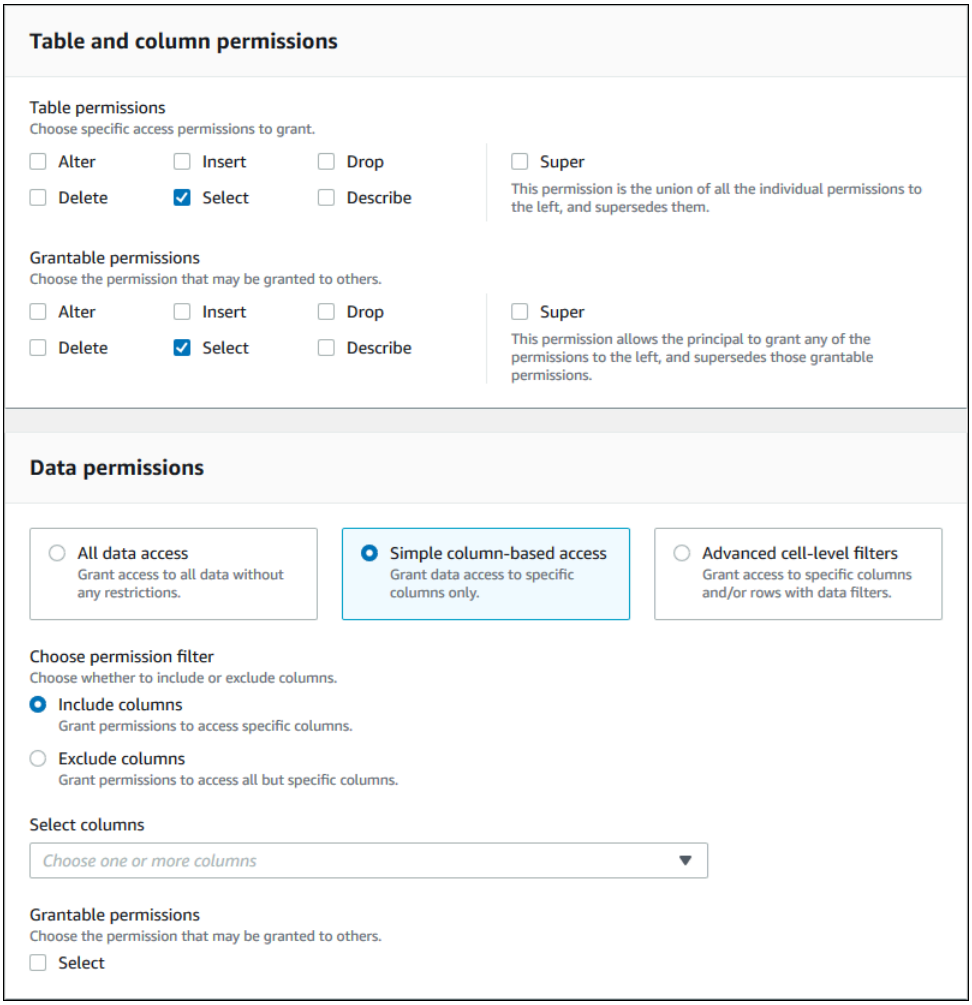
5. Scegli Concessione.
6. Specificare l'autorizzazione Seleziona con filtraggio dei dati

   Seleziona l'autorizzazione Seleziona. Non selezionare altre autorizzazioni.

   La sezione Autorizzazioni per i dati viene visualizzata sotto la sezione Autorizzazioni per tabelle e colonne.
7. Esegui una di queste operazioni:

   - Applica solo un semplice filtraggio delle colonne.

     1. Scegli Accesso semplice basato su colonne.



2. Scegli se includere o escludere le colonne, quindi scegli le colonne da includere o escludere.

   Quando si concedono le autorizzazioni a un AWS account o a un'organizzazione esterni, sono supportati solo gli elenchi di inclusione.

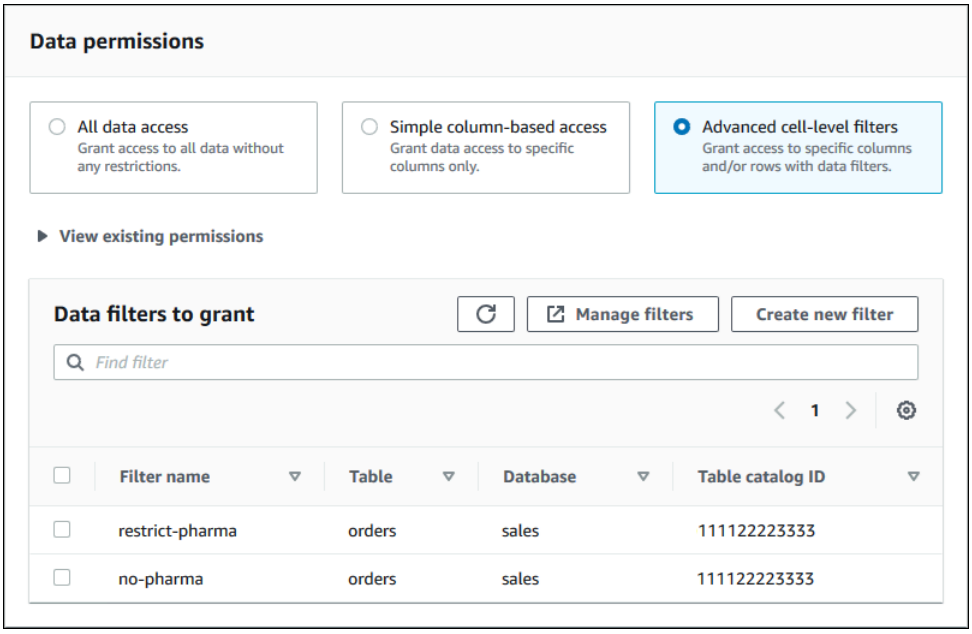
3. (Facoltativo) In Autorizzazioni concedibili, attiva l'opzione di concessione per l'autorizzazione Seleziona.

   Se includi l'opzione di concessione, il destinatario della concessione può concedere le autorizzazioni solo nelle colonne che gli concedi.

> **Note**
>
> Puoi anche applicare il filtro delle colonne solo creando un filtro dati che specifichi un filtro di colonna e specifichi tutte le righe come filtro di riga. Tuttavia, ciò richiede ulteriori passaggi.

- Applica il filtraggio di colonne, righe o celle.

  1. Scegli Filtri avanzati a livello di cella.



  2. (Facoltativo) Espandi Visualizza le autorizzazioni esistenti.
  3. (Facoltativo) Scegli Crea nuovo filtro.
  4. (Facoltativo) Per visualizzare i dettagli dei filtri elencati o per creare nuovi filtri o eliminare filtri esistenti, scegli Gestisci filtri.

     La pagina Filtri dati si apre in una nuova finestra del browser.

     Al termine della pagina Filtri dati, torna alla pagina Concedi autorizzazioni e, se necessario, aggiorna la pagina per visualizzare i nuovi filtri di dati che hai creato.
  5. Seleziona uno o più filtri di dati da applicare alla concessione.

> **Note**
>
> Se non ci sono filtri di dati nell'elenco, significa che non sono stati creati filtri di dati per la tabella selezionata.

8. Scegli Concessione.

AWS CLI

È possibile concedere le autorizzazioni per la tabella utilizzando il metodo di risorsa denominato e il AWS Command Line Interface (AWS CLI).

Per concedere i permessi alle tabelle utilizzando il AWS CLI

- Eseguite un `grant-permissions` comando e specificate una tabella come risorsa.

Example — Concessione su una singola tabella, senza filtri

L'esempio seguente concede `SELECT` e `ALTER` all'utente `datalake_user1` nell'AWSaccount 1111-2222-3333 sulla tabella del database. `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 concedi l'`ALTER`autorizzazione su una tabella i cui dati sottostanti si trovano in una posizione registrata, assicurati di concedere anche ai principali i permessi di localizzazione dei dati sulla posizione. Per ulteriori informazioni, consulta Concessione delle autorizzazioni per la localizzazione dei dati.

Example — Concedi su tutte le tabelle con l'opzione Grant: nessun filtro

L'esempio successivo concede `SELECT` con l'opzione grant su tutte le tabelle del database. `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 con un semplice filtraggio delle colonne

Questo esempio successivo concede autorizzazioni `SELECT` su un sottoinsieme di colonne della tabella. `persons` Utilizza un semplice filtraggio delle colonne.

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "SELECT" --resource '{ "TableWithColumns": {"DatabaseName":"hr",
 "Name":"persons", "ColumnNames":["family_name", "given_name", "gender"]}}'
```

Example — Concedi con un filtro dati

Questo esempio concede `SELECT` sulla `orders` tabella e applica il filtro `restrict-pharma` dati.

```
aws lakeformation grant-permissions --cli-input-json file://grant-params.json
```

Di seguito è riportato il contenuto del file`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"]
}
```

Consulta anche

- Panoramica delle autorizzazioni di Lake Formation
- Filtraggio dei dati e sicurezza a livello di cella in Lake Formation
- Riferimento ai personaggi di Lake Formation e alle autorizzazioni IAM
- Concessione delle autorizzazioni per i collegamenti alle risorse
- Accesso e visualizzazione di tabelle e database condivisi del Data Catalog

Concessione delle autorizzazioni sulle viste utilizzando il metodo di risorsa denominato

I passaggi seguenti spiegano come concedere le autorizzazioni sulle viste utilizzando il metodo della risorsa denominata e la pagina Concedi le autorizzazioni del data lake. La pagina è suddivisa nelle seguenti sezioni:

- Principi: gli utenti, i ruoli, gli utenti e i gruppi di IAM Identity CenterAccount AWS, le organizzazioni o le unità organizzative a cui concedere le autorizzazioni.
- LF-tags o risorse di catalogo: database, tabelle, viste o link alle risorse su cui concedere le autorizzazioni.
- Autorizzazioni: le autorizzazioni del data lake da concedere.

Apri la pagina Concedi le autorizzazioni del data lake

1. Apri la AWS Lake Formation console all'indirizzo https://console.aws.amazon.com/lakeformation/ e accedi come amministratore del data lake, creatore del database o utente IAM con autorizzazioni Grantable sul database.
2. Esegui una di queste operazioni:
   - Nel riquadro di navigazione, in Autorizzazioni, scegli Autorizzazioni Data Lake. Quindi scegli Concedi.
   - Nel riquadro di navigazione, scegli Visualizzazioni in Data Catalog. Quindi, nella pagina Visualizzazioni, scegli una visualizzazione e dal menu Azioni, in Autorizzazioni, scegli Concedi.

> **Note**
>
> Puoi concedere le autorizzazioni per una vista tramite il relativo link alla risorsa. A tale scopo, nella pagina Visualizzazioni, scegli un collegamento a una risorsa e nel menu Azioni scegli Concedi all'obiettivo. Per ulteriori informazioni, consulta Come funzionano i link alle risorse in Lake Formation.

Specificate i principi

Nella sezione Principi, scegli un tipo principale, quindi specifica i principali a cui concedere le autorizzazioni.

Utenti e ruoli IAM

Scegli uno o più utenti o ruoli dall'elenco degli utenti e dei ruoli IAM.

IAM Identity Center

Scegli uno o più utenti o gruppi dall'elenco Utenti e gruppi.

Utenti e gruppi SAML

Per QuickSight utenti e gruppi SAML e Amazon, inserisci uno o più Amazon Resource Names (ARN) per utenti o gruppi federati tramite SAML o ARN per utenti o gruppi Amazon. QuickSight Premi Invio dopo ogni ARN.

Per informazioni su come costruire gli ARN, vedere. Comandi di concessione e AWS CLI revoca di Lake Formation

> **Note**
>
> L'integrazione di Lake Formation con Amazon QuickSight è supportata solo per Amazon QuickSight Enterprise Edition.

Account esterni

Per Account AWS, AWS organizzazione o Principal IAM inserisci uno o più ID AWS account, ID organizzazione, ID unità organizzative o ARN validi per l'utente o il ruolo IAM. Premi Invio dopo ogni ID.

Un ID dell'organizzazione è composto da «o-» seguito da 10-32 lettere o cifre minuscole.

L'ID di un'unità organizzativa inizia con «ou-» seguito da 4—32 lettere o cifre minuscole (l'ID della radice che contiene l'unità organizzativa). Questa stringa è seguita da un secondo trattino «-» e da 8 a 32 lettere o cifre minuscole aggiuntive.

Vedi anche

- Accesso e visualizzazione di tabelle e database condivisi del Data Catalog

Specificate le viste

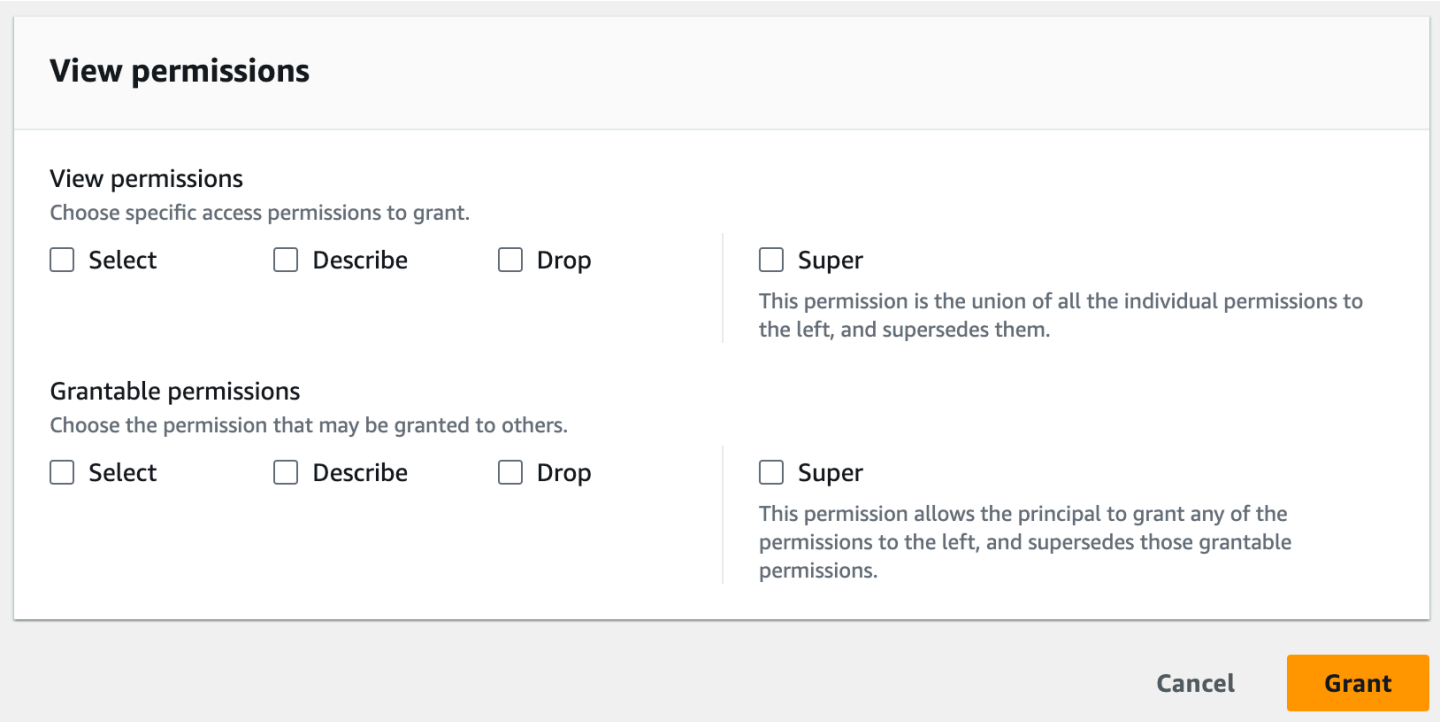
Nella sezione LF-tags o risorse del catalogo, scegliete una o più viste a cui concedere le autorizzazioni.

1. Scegliete Named data catalog resources.
2. Scegliete una o più viste dall'elenco Visualizzazioni. Puoi anche scegliere uno o più database, tabelle e/o filtri di dati.

   La concessione delle autorizzazioni per il data lake all'`All views`interno di un database comporterà che il beneficiario disponga delle autorizzazioni su tutte le tabelle e le viste all'interno del database.

Specificare le autorizzazioni

Nella sezione Autorizzazioni, seleziona autorizzazioni e autorizzazioni concedibili.



1. In Visualizza autorizzazioni, seleziona una o più autorizzazioni da concedere.
2. (Facoltativo) In Autorizzazioni concedibili, seleziona le autorizzazioni che il destinatario della sovvenzione può concedere ad altri responsabili del proprio territorio. Account AWS Questa opzione non è supportata quando si concedono autorizzazioni a un responsabile IAM da un account esterno.
3. Scegli Concessione.

Vedi anche

- Riferimento alle autorizzazioni di Lake Formation
- Concessione delle autorizzazioni su un database o una tabella condivisa con il tuo account

Controllo degli accessi basato su tag Lake Formation

Il controllo degli accessi basato su tag di Lake Formation (LF-TBAC) è una strategia di autorizzazione che definisce le autorizzazioni in base agli attributi. In Lake Formation, questi attributi sono chiamati LF-tag. È possibile allegare tag LF alle risorse del Data Catalog e concedere autorizzazioni ai responsabili di Lake Formation su tali risorse utilizzando questi tag LF. Lake Formation consente

operazioni su tali risorse quando il valore del tag del principale corrisponde al valore del tag della risorsa. LF-TBAC è utile in ambienti in rapida crescita e aiuta in situazioni in cui la gestione delle politiche diventa complicata.

LF-TBAC è il metodo consigliato da utilizzare per concedere le autorizzazioni di Lake Formation quando è presente un numero elevato di risorse del Data Catalog. LF-TBAC è più scalabile rispetto al metodo Named Resource e richiede un minor sovraccarico di gestione delle autorizzazioni.

> **Note**
>
> I tag IAM non sono gli stessi dei tag LF. Questi tag non sono intercambiabili. I tag LF vengono utilizzati per concedere i permessi di Lake Formation e i tag IAM vengono utilizzati per definire le politiche IAM.

Come funziona il controllo degli accessi basato su tag Lake Formation

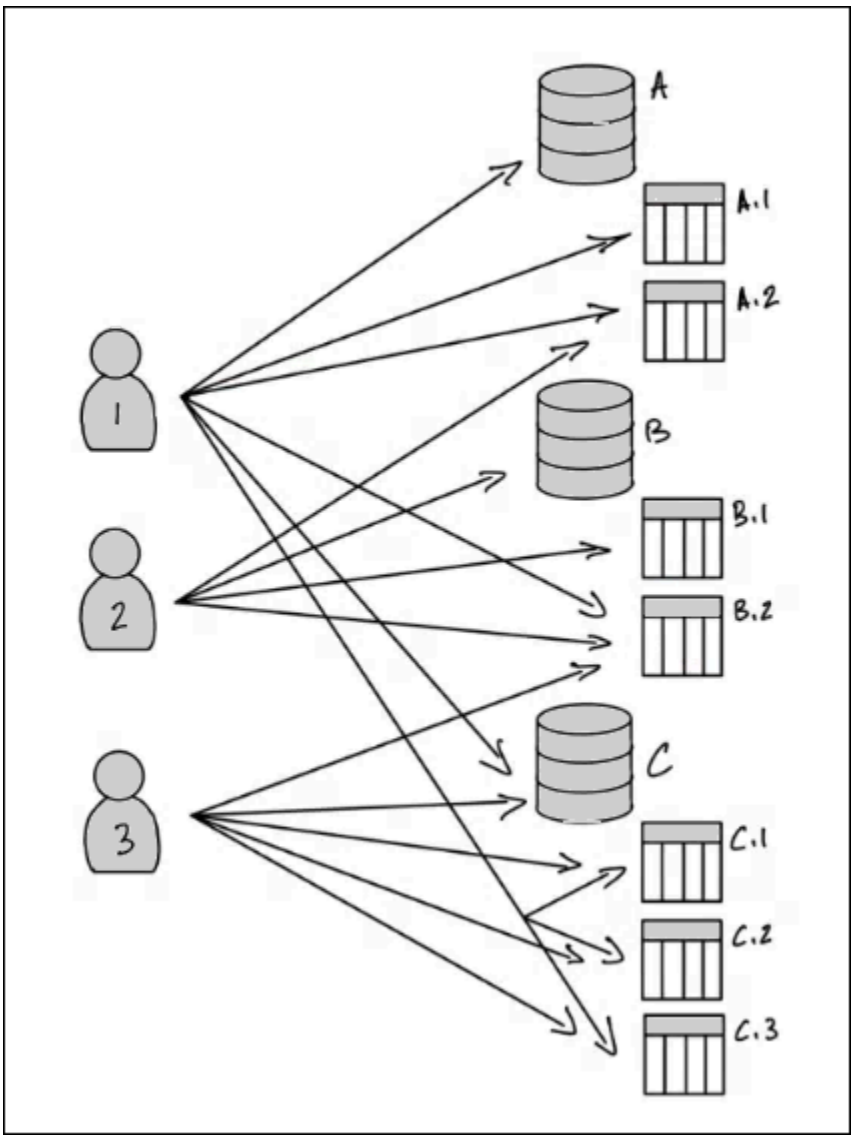
Ogni tag LF è una coppia chiave-valore, ad esempio o. `department=sales` `classification=restricted` Una chiave può avere più valori definiti, ad esempio. `department=sales,marketing,engineering,finance`

Per utilizzare il metodo LF-TBAC, gli amministratori del data lake e i data engineer eseguono le seguenti attività.

Attività	Dettagli dell'attività
1. Definire le proprietà e le relazioni dei tag LF.	-
2. Crea i creatori di tag LF in Lake Formation.	Aggiungere creatori di tag LF
3. Crea il tag LF in Lake Formation.	Creazione di tag LF
4. Assegna i tag LF alle risorse del Data Catalog.	Assegnazione di tag LF alle risorse del Data Catalog

Attività	Dettagli dell'attività
5. Concedi le autorizzazioni ad altri principali per assegnare i tag LF alle risorse, opzionalmente con l'opzione grant.	Concessione, revoca ed elenco delle autorizzazioni per i valori LF-Tag
6. Concedi le espressioni dei tag LF ai principali, opzionalmente con l'opzione grant.	Concessione delle autorizzazioni per il data lake utilizzando il metodo LF-TBAC
7. (Consigliato) Dopo aver verificato che i principali abbiano accesso alle risorse corrette tramite il metodo LF-TBAC, revocate le autorizzazioni concesse utilizzando il metodo della risorsa denominata.	-

Si consideri il caso in cui è necessario concedere le autorizzazioni a tre principali su tre database e sette tabelle.



Per ottenere le autorizzazioni indicate nel diagramma precedente utilizzando il metodo named resource, è necessario concedere 17 concessioni, come segue (in pseudo-codice).

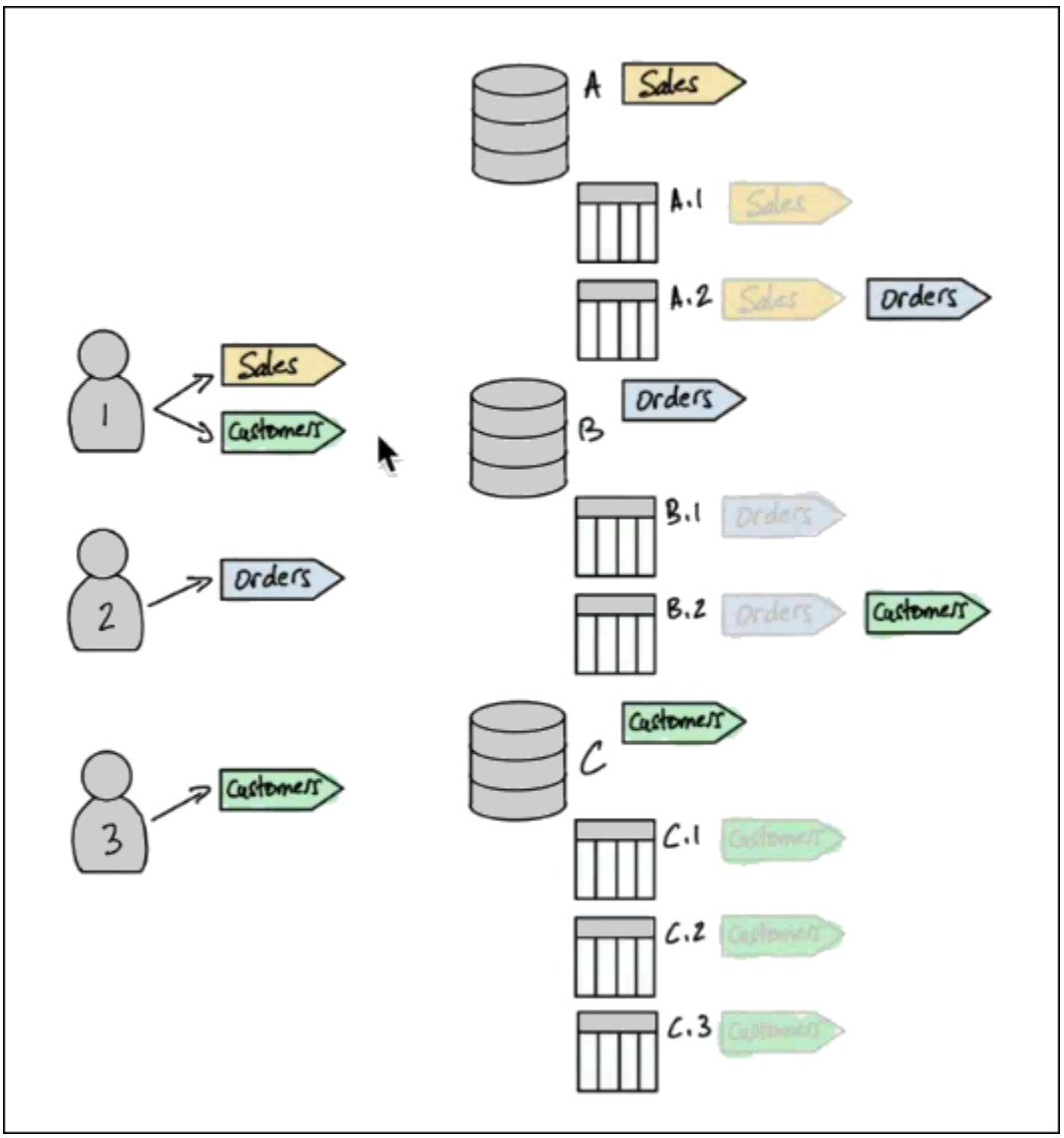
```
GRANT CREATE_TABLE ON Database A TO PRINCIPAL 1
GRANT SELECT, INSERT ON Table A.1 TO PRINCIPAL 1
GRANT SELECT, INSERT ON Table A.2 TO PRINCIPAL 1
GRANT SELECT, INSERT ON Table B.2 TO PRINCIPAL 1
...
GRANT SELECT, INSERT ON Table A.2 TO PRINCIPAL 2
GRANT CREATE_TABLE ON Database B TO PRINCIPAL 2
...
GRANT SELECT, INSERT ON Table C.3 TO PRINCIPAL 3
```

Considerate ora come concedereste le autorizzazioni utilizzando LF-TBAC. Il diagramma seguente indica che sono stati assegnati tag LF a database e tabelle e che sono state concesse le autorizzazioni sui tag LF ai principali.

In questo esempio, i tag LF rappresentano aree del data lake che contengono analisi per diversi moduli di una suite di applicazioni ERP (Enterprise Resource Planning). È possibile controllare l'accesso ai dati di analisi per i vari moduli. Tutti i tag LF hanno la chiave `module` e i valori possibili `SalesOrders`, e. `Customers` Un esempio di tag LF ha il seguente aspetto:

```
module=Sales
```

Il diagramma mostra solo i valori del tag LF.



Assegnazione di tag alle risorse e all'ereditarietà di Data Catalog

Le tabelle ereditano i tag LF dai database e le colonne ereditano i tag LF dalle tabelle. I valori ereditati possono essere sovrascritti. Nel diagramma precedente, i tag LF oscurati vengono ereditati.

A causa dell'ereditarietà, l'amministratore del data lake deve effettuare solo le seguenti cinque assegnazioni di tag LF alle risorse (in pseudo-codice).

```
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
```

Contributi di tag ai committenti

Dopo aver assegnato i tag LF ai database e alle tabelle, l'amministratore del data lake deve concedere solo quattro tag LF ai principali, come segue (in pseudo-codice).

```
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
```

Ora, un principale con il tag `module=Sales` LF può accedere alle risorse del Data Catalog con il tag LF (ad esempio, il database A), un principale con il `module=Sales` tag LF può accedere alle risorse con il tag LF e così via. `module=Customers` `module=Customers`

I comandi di concessione precedenti sono incompleti. Questo perché, sebbene indichino tramite LF-Tags le risorse del Data Catalog su cui i responsabili hanno i permessi, non indicano esattamente quali `SELECT` permessi di Lake Formation (ad esempio`ALTER`) i principali hanno su tali risorse. Pertanto, i seguenti comandi in pseudo-codice sono una rappresentazione più accurata del modo in cui le autorizzazioni di Lake Formation vengono concesse alle risorse del Data Catalog tramite i tag LF.

```
GRANT (CREATE_TABLE ON DATABASES) ON TAGS module=Sales TO Principal 1
GRANT (SELECT, INSERT ON TABLES)  ON TAGS module=Sales TO Principal 1
GRANT (CREATE_TABLE ON DATABASES) ON TAGS module=Customers TO Principal 1
GRANT (SELECT, INSERT ON TABLES)  ON TAGS module=Customers TO Principal 1
GRANT (CREATE_TABLE ON DATABASES) ON TAGS module=Orders TO Principal 2
GRANT (SELECT, INSERT ON TABLES)  ON TAGS module=Orders TO Principal 2
GRANT (CREATE_TABLE ON DATABASES) ON TAGS module=Customers TO Principal 3
```

```
GRANT (SELECT, INSERT ON TABLES)  ON TAGS module=Customers TO Principal 3
```

Mettere insieme: autorizzazioni risultanti sulle risorse

Dati i tag LF assegnati ai database e alle tabelle nel diagramma precedente e i tag LF concessi ai principali del diagramma, la tabella seguente elenca i permessi di Lake Formation che i principali hanno sui database e sulle tabelle.

Principale	Autorizzazioni concesse tramite LF-Tags
Principale 1	• `CREATE_TABLE` sul database A • `SELECT`, `INSERT` nella tabella A.1 • `SELECT`, `INSERT` nella tabella B.2 • `CREATE_TABLE` nel database C • `SELECT`, `INSERT` nella tabella C.1 • `SELECT`, `INSERT` nella tabella C.2 • `SELECT`, `INSERT` nella tabella C.3
Principal 2	• `SELECT`, `INSERT` nella tabella A.2 • `CREATE_TABLE` nel database B • `SELECT`, `INSERT` nella tabella B.1 • `SELECT`, `INSERT` nella tabella B.2
Principal 3	• `SELECT`, `INSERT` nella tabella B.2 • `CREATE_TABLE` nel database C • `SELECT`, `INSERT` nella tabella C.1 • `SELECT`, `INSERT` nella tabella C.2 • `SELECT`, `INSERT` nella tabella C.3

Conclusione

In questo semplice esempio, utilizzando cinque operazioni di assegnazione e otto operazioni di concessione, l'amministratore del data lake è stato in grado di specificare 17 autorizzazioni. Quando ci sono decine di database e centinaia di tabelle, il vantaggio del metodo LF-TBAC rispetto al metodo

delle risorse denominate diventa evidente. Nel caso ipotetico della necessità di concedere a ogni risorsa l'accesso principale, e `n(P)` dov'è il numero di principali e il numero di risorse: `n(R)`

- Con il metodo Named Resource, il numero di sovvenzioni richieste è ×. `n(P) n(R)`
- Con il metodo LF-TBAC, utilizzando un singolo tag LF, il totale del numero di sovvenzioni ai committenti e di assegnazioni alle risorse è +. `n(P) n(R)`

> Consulta anche
>
> - Gestione dei tag LF per il controllo dell'accesso ai metadati
> - Concessione delle autorizzazioni per il data lake utilizzando il metodo LF-TBAC

Argomenti

Gestione dei tag LF per il controllo dell'accesso ai metadati

Per utilizzare il metodo di controllo degli accessi basato su tag Lake Formation (LF-TBAC) per proteggere le risorse del Data Catalog (database, tabelle e colonne), devi creare tag LF, assegnarli alle risorse e concedere le autorizzazioni LF-Tag ai principali.

Prima di poter assegnare i tag LF alle risorse del Data Catalog o concedere le autorizzazioni ai principali, è necessario definire i tag LF. Solo un amministratore del data lake o un responsabile con i permessi di creazione di LF-Tag può creare LF-Tag.

Creatori di tag LF

LF-Tag creator è un amministratore non amministratore che dispone delle autorizzazioni per creare e gestire i tag LF. Gli amministratori di Data Lake possono aggiungere creatori di tag LF utilizzando la console di Lake Formation o la CLI. I creatori di LF-Tag hanno i permessi impliciti di Lake Formation per aggiornare ed eliminare i tag LF, per assegnare i tag LF alle risorse e per concedere i permessi dei tag LF e i permessi per i valori dei tag LF ad altri principali.

Con i ruoli di creatore di LF-Tag, gli amministratori del data lake possono delegare attività di gestione dei tag come la creazione e l'aggiornamento di chiavi e valori dei tag a responsabili non

amministratori. Gli amministratori di Data Lake possono anche concedere autorizzazioni valide ai creatori di LF-Tag. `Create LF-Tag` Quindi, il creatore del tag LF può concedere il permesso di creare tag LF ad altri responsabili.

È possibile concedere due tipi di autorizzazioni sui tag LF:

- Autorizzazioni LF-Tag -, e. `Create LF-Tag` `Alter` `Drop` Queste autorizzazioni sono necessarie per creare, aggiornare ed eliminare i tag LF.

  Gli amministratori di Data Lake e i creatori di LF-Tag dispongono implicitamente di queste autorizzazioni sui tag LF che creano e possono concedere queste autorizzazioni esplicitamente ai responsabili della gestione dei tag nel data lake.
- `Grant with LF-Tag expressions`Autorizzazioni `Assign` della `Describe` coppia chiave-valore LF-Tag -,, e. Queste autorizzazioni sono necessarie per assegnare i tag LF ai database, alle tabelle e alle colonne del Data Catalog e per concedere le autorizzazioni sulle risorse ai responsabili che utilizzano il controllo degli accessi basato su tag Lake Formation. I creatori di LF-Tag ricevono implicitamente queste autorizzazioni durante la creazione di LF-Tag.

Dopo aver ricevuto l'`Create LF-Tag`autorizzazione e aver creato con successo i tag LF, il creatore di LF-Tag può assegnare i tag LF alle risorse e concedere i permessi LF-Tag (`Create LF-Tag`,, e) ad altri titolari non amministrativi per gestire i tag nel data lake. `Alter` `Drop` Puoi gestire i tag LF utilizzando la console Lake Formation, l'API o AWS Command Line Interface ()AWS CLI.

> **Note**
>
> Gli amministratori di Data Lake dispongono delle autorizzazioni implicite di Lake Formation per creare, aggiornare ed eliminare i tag LF, assegnare i tag LF alle risorse e concedere i permessi LF-Tag ai principali.

Per le migliori pratiche e considerazioni, consulta Buone pratiche e considerazioni per il controllo degli accessi basato su tag Lake Formation

Argomenti

- Aggiungere creatori di tag LF
- Creazione di tag LF
- Aggiornamento dei tag LF

- Eliminazione dei tag LF
- Elenco dei tag LF
- Assegnazione di tag LF alle risorse del Data Catalog
- Visualizzazione dei tag LF assegnati a una risorsa
- Visualizzazione delle risorse a cui è assegnato un LF-Tag
- Ciclo di vita di un tag LF
- Confronto tra il controllo degli accessi basato su tag Lake Formation e il controllo degli accessi basato sugli attributi IAM

> **Consulta anche**
>
> - Concessione, revoca ed elenco delle autorizzazioni per i valori LF-Tag
> - Concessione delle autorizzazioni per il data lake utilizzando il metodo LF-TBAC
> - Controllo degli accessi basato su tag Lake Formation

Aggiungere creatori di tag LF

Per impostazione predefinita, gli amministratori del data lake possono creare, aggiornare ed eliminare i tag LF, assegnare tag alle risorse del Data Catalog e concedere le autorizzazioni relative ai tag ai principali. Se desideri delegare le operazioni di creazione e gestione dei tag a responsabili non amministratori, l'amministratore del data lake può creare ruoli di creatore di tag LF e concedere l'autorizzazione a Lake Formation ai ruoli. `Create LF-Tag` Con l'`Create LF-Tag`autorizzazione concessa, i creatori di LF-Tag possono delegare le attività di creazione e manutenzione dei tag ad altri responsabili non amministrativi.

> **Note**
>
> Le concessioni di autorizzazioni per più account possono includere solo autorizzazioni e. `Describe` `Associate` Non puoi concedere`Create LF-Tag`, `Drop``Alter`, e `Grant with LFTag expressions` autorizzazioni ai responsabili di un altro account.

Argomenti

- Autorizzazioni IAM necessarie per creare tag LF
- Aggiungi creatori di tag LF

> **Consulta anche**
>
> - Concessione, revoca ed elenco delle autorizzazioni per i valori LF-Tag
> - Concessione delle autorizzazioni per il data lake utilizzando il metodo LF-TBAC
> - Controllo degli accessi basato su tag Lake Formation

Autorizzazioni IAM necessarie per creare tag LF

È necessario configurare le autorizzazioni per consentire a un responsabile di Lake Formation di creare tag LF. Aggiungi la seguente dichiarazione alla politica delle autorizzazioni per il principale che deve essere un creatore di LF-Tag.

> **Note**
>
> Sebbene gli amministratori dei data lake abbiano le autorizzazioni implicite di Lake Formation per creare, aggiornare ed eliminare i tag LF, assegnare i tag LF alle risorse e concedere i tag LF ai principali, gli amministratori del data lake necessitano anche delle seguenti autorizzazioni IAM.

Per ulteriori informazioni, consulta Riferimento ai personaggi di Lake Formation e alle autorizzazioni IAM.

```
{
"Sid": "Transformational",
"Effect": "Allow",
    "Action": [
        "lakeformation:AddLFTagsToResource",
        "lakeformation:RemoveLFTagsFromResource",
        "lakeformation:GetResourceLFTags",
        "lakeformation:ListLFTags",
        "lakeformation:CreateLFTag",
        "lakeformation:GetLFTag",
        "lakeformation:UpdateLFTag",
```

```
        "lakeformation:DeleteLFTag",
        "lakeformation:SearchTablesByLFTags",
        "lakeformation:SearchDatabasesByLFTags"
    ]
}
```

I principali che assegnano i tag LF alle risorse e concedono i tag LF ai principali devono avere le stesse autorizzazioni, ad eccezione delle autorizzazioni, e. `CreateLFTag` `UpdateLFTag` `DeleteLFTag`

Aggiungi creatori di tag LF

Un creatore di tag LF può creare un tag LF, aggiornare la chiave e i valori dei tag, eliminare tag, associare tag alle risorse del Data Catalog e concedere le autorizzazioni sulle risorse del Data Catalog ai principali utilizzando il metodo LF-TBAC. Il creatore di LF-Tag può anche concedere queste autorizzazioni ai principali.

È possibile creare ruoli di creatore di LF-Tag utilizzando la AWS Lake Formation console, l'API o (). AWS Command Line Interface AWS CLI

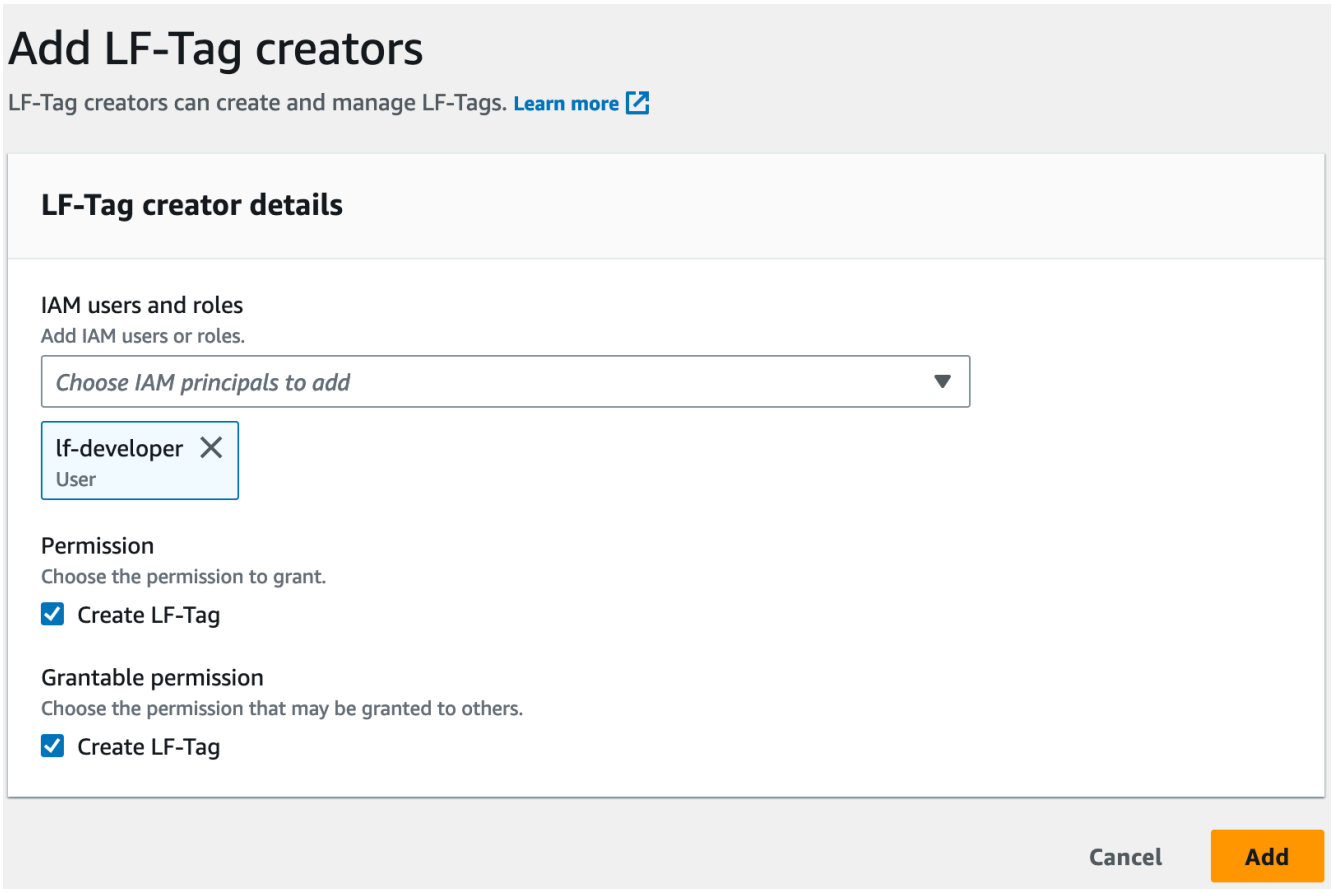
console

Per aggiungere un creatore di tag LF

1. Aprire la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.

   Accedi come amministratore del datalake.
2. Nel pannello di navigazione, in Autorizzazioni, scegli LF-tags e permessi.

   Nella pagina LF-Tag e permessi, scegliete la sezione LF-Tag Creators e scegliete Aggiungi creatori LF-Tag.



3. Nella pagina Aggiungi creatori di LF-Tag, scegli un ruolo o un utente IAM con le autorizzazioni necessarie per creare LF-Tag.
4. Abilita `Create LF-Tag` la casella di controllo delle autorizzazioni.
5. (Facoltativo) Per consentire ai responsabili selezionati di concedere l'`Create LF-Tag`autorizzazione ai mandanti, scegli Autorizzazione `Create LF-Tag` concedibile.
6. Scegli Aggiungi.

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"
    ]
}
```

Le seguenti sono le autorizzazioni disponibili per il ruolo di creatore di LF-Tag:

Autorizzazione	Descrizione
`Drop`	Un principale con questa autorizzazione su un LF-Tag può eliminare un LF-Tag dal data lake. Il principale ottiene l'`Describe`autorizzazione implicita su tutti i valori dei tag di una risorsa LF-Tag.
`Alter`	Un principale con questa autorizzazione su un tag LF può aggiunger e o rimuovere un valore di tag da un tag LF. Il principale ottiene l'`Alter`autorizzazione implicita su tutti i valori dei tag di un tag LF.
`Describe`	Un principale con questa autorizzazione su un tag LF può visualizzare il tag LF e i suoi valori quando assegna tag LF alle risorse o concede autorizzazioni sui tag LF. È possibile concedere su tutti i valori chiave o su valori specifici. `Describe`
`Associate`	Un principale con questa autorizzazione su un tag LF può assegnare il tag LF a una risorsa del catalogo dati. `Associate` `Describe`Concedere concessioni implicite.
`Grant with LF-Tag expression`	Un principale con questa autorizzazione su un LF-Tag può concedere autorizzazioni sulle risorse di un Data Catalog utilizzando la chiave e i valori LF-Tag. `Grant with LF-Tag expression` La concessione implicita di concessioni. `Describe`

Queste autorizzazioni sono concesse. Un preside a cui sono state concesse queste autorizzazioni con l'opzione di concessione può concederle ad altri committenti.

Creazione di tag LF

Tutti i tag LF devono essere definiti in Lake Formation prima di poter essere utilizzati. Un tag LF è composto da una chiave e da uno o più valori possibili per la chiave.

Dopo che l'amministratore del data lake ha configurato le autorizzazioni IAM e le autorizzazioni Lake Formation richieste per il ruolo di creatore di LF-Tag, il responsabile può creare un LF-Tag. Il creatore del tag LF ottiene l'autorizzazione implicita ad aggiornare o rimuovere qualsiasi valore di tag dal tag LF e ad eliminare il tag LF.

È possibile creare tag LF utilizzando la AWS Lake Formation console, l'API o (). AWS Command Line Interface AWS CLI

Console

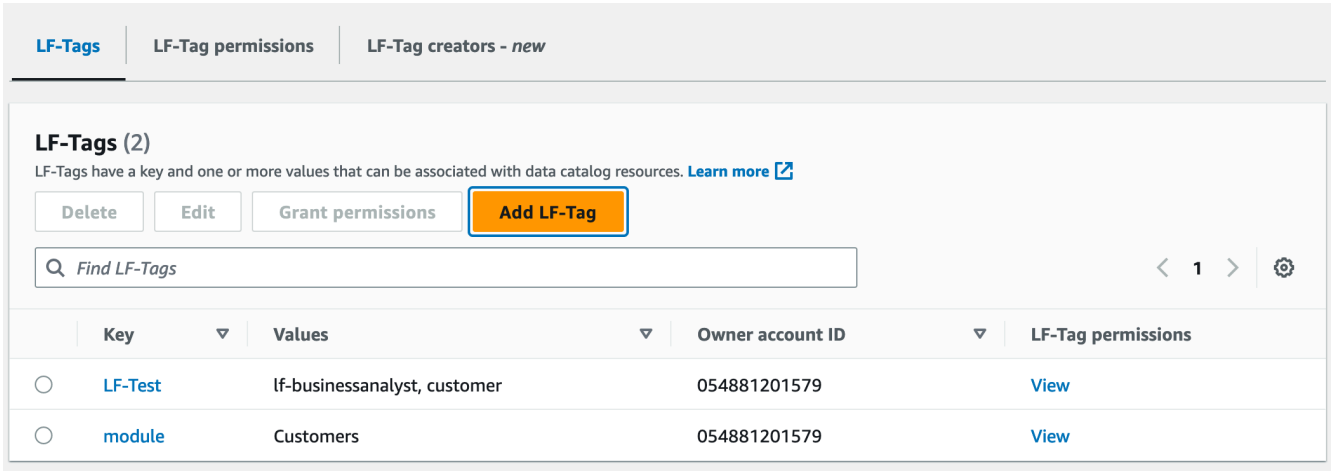
Per creare un tag LF

1. Aprire la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.

   Accedi come principale con i permessi di creazione di LF-Tag o come amministratore del data lake.

2. Nel pannello di navigazione, sotto LF-tags e permessi, scegli LF-Tags.

   Viene visualizzata la pagina LF-Tags.



3. Scegli Aggiungi tag LF.
4. Nella finestra di dialogo Aggiungi LF-tag, inserite una chiave e uno o più valori.

Ogni chiave deve avere almeno un valore. Per inserire più valori, inserisci un elenco delimitato da virgole e premi Invio oppure inserisci un valore alla volta e scegli Aggiungi dopo ognuno di essi. Il numero massimo di valori consentito è 1000.

5. Selezionare Aggiungi tag.

AWS CLI

Per creare un tag LF

- Inserire un comando`create-lf-tag`.

  L'esempio seguente crea un tag LF con chiave `module` e valori `Customers` e. `Orders`

```
aws lakeformation create-lf-tag --tag-key module --tag-values Customers Orders
```

In qualità di creatore di tag, il principale ottiene l'`Alter`autorizzazione per questo tag LF e può aggiornare o rimuovere qualsiasi valore di tag da questo tag LF. Il responsabile del creatore di LF-Tag può anche concedere `Alter` il permesso a un altro principale di aggiornare e rimuovere i valori dei tag su questo LF-Tag.

Aggiornamento dei tag LF

Aggiorna un tag LF per il quale hai l'`Alter`autorizzazione aggiungendo o eliminando i valori chiave consentiti. Non è possibile modificare la chiave LF-Tag. Per cambiare la chiave, elimina il tag LF e aggiungine uno con la chiave richiesta. Oltre all'`Alter`autorizzazione, è necessaria anche l'autorizzazione `lakeformation:UpdateLFTag` IAM per aggiornare i valori.

Quando si elimina un valore del tag LF, non viene eseguito alcun controllo per verificare la presenza di tale valore in alcuna risorsa del Data Catalog. Se il valore del tag LF eliminato è associato a una risorsa, non è più visibile per la risorsa e tutti i principali a cui sono state concesse le autorizzazioni per quella coppia chiave-valore non dispongono più delle autorizzazioni.

Prima di eliminare un valore del tag LF, è possibile utilizzare facoltativamente il `remove-lf-tags-from-resource`comando per rimuovere il tag LF dalle risorse del Catalogo dati che hanno il valore che si desidera eliminare, quindi rietichettare la risorsa con i valori che si desidera conservare.

Solo gli amministratori del data lake, il creatore del tag LF e i responsabili che dispongono delle autorizzazioni per l'LF-Tag possono aggiornare un LF-Tag. `Alter`

È possibile aggiornare un LF-Tag utilizzando la console, l'API o il (). AWS Lake Formation AWS Command Line Interface AWS CLI

Console

Per aggiornare un LF-tag (console)

1. Aprire la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.

   Accedi come amministratore del data lake, creatore di LF-Tag o titolare con l'`Alter`autorizzazione per l'LF-Tag.
2. Nel pannello di navigazione, sotto LF-Tag e autorizzazioni, scegli LF-Tags.
3. Nella pagina LF-Tags, selezionate un LF-Tag, quindi scegliete Modifica.
4. Nella finestra di dialogo Modifica LF-Tag, aggiungete o rimuovete i valori del tag LF.

   Per aggiungere più valori, nel campo Valori, inserite un elenco delimitato da virgole e premete Invio, oppure inserite un valore alla volta o scegliete Aggiungi dopo ognuno di essi.
5. Selezionare Salva.

AWS CLI

Per aggiornare un tag LF () AWS CLI

- Immettete un comando`update-lf-tag`. Fornite uno o entrambi i seguenti argomenti:
  - `--tag-values-to-add`
  - `--tag-values-to-delete`

Example

L'esempio seguente sostituisce il valore `vp` con il valore della chiave `vice-president` LF-Tag. `level`

```
aws lakeformation update-lf-tag --tag-key level --tag-values-to-add vice-president
--tag-values-to-delete vp
```

Eliminazione dei tag LF

È possibile eliminare i tag LF che non sono più in uso. Non viene eseguito alcun controllo per la presenza del tag LF su una risorsa del Data Catalog. Se il tag LF eliminato è associato a una risorsa, non è più visibile per la risorsa e tutti i principali a cui sono state concesse le autorizzazioni su quel tag LF non dispongono più delle autorizzazioni.

Prima di eliminare un tag LF, è possibile utilizzare facoltativamente il comando per rimuovere il tag LF da tutte le risorse. `remove-lf-tags-from-resource`

Solo gli amministratori del data lake, il creatore del tag LF o un responsabile che dispone dell'autorizzazione per l'LF-Tag possono eliminare un LF-Tag. `Drop` Oltre all'autorizzazione, il principale necessita anche dell'`Drop`autorizzazione IAM per eliminare un LF-Tag. `lakeformation:DeleteLFTag`

È possibile eliminare un tag LF utilizzando la AWS Lake Formation console, l'API o (). AWS Command Line Interface AWS CLI

Console

Per eliminare un LF-tag (console)

1. Aprire la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.

   Accedi come amministratore del data lake.
2. Nel pannello di navigazione, sotto LF-tags e permessi, scegli LF-Tags.
3. Nella pagina LF-Tags, selezionate un LF-Tag, quindi scegliete Elimina.
4. Nell'ambiente Delete tag? nella finestra di dialogo, per confermare l'eliminazione, inserite il valore della chiave LF-Tag nel campo designato, quindi scegliete Elimina.

AWS CLI

Per eliminare un tag LF () AWS CLI

- Immettete un comando`delete-lf-tag`. Fornisci la chiave del tag LF da eliminare.

  Example

  L'esempio seguente elimina il tag LF con la chiave. `region`

```
aws lakeformation delete-lf-tag --tag-key region
```

Elenco dei tag LF

Puoi elencare i tag LF su cui hai i permessi o. `Describe` `Associate` I valori elencati con ogni chiave LF-Tag sono i valori per i quali avete i permessi.

LF-Tag creator ha i permessi impliciti per vedere i tag LF che ha creato.

Gli amministratori di Data Lake possono vedere tutti i tag LF definiti nell'AWSaccount locale e tutti i tag LF per i quali sono state concesse le `Associate` autorizzazioni `Describe` e all'account locale da account esterni. L'amministratore del data lake può vedere tutti i valori di tutti i tag LF.

È possibile elencare i tag LF utilizzando la AWS Lake Formation console, l'API o (). AWS Command Line Interface AWS CLI

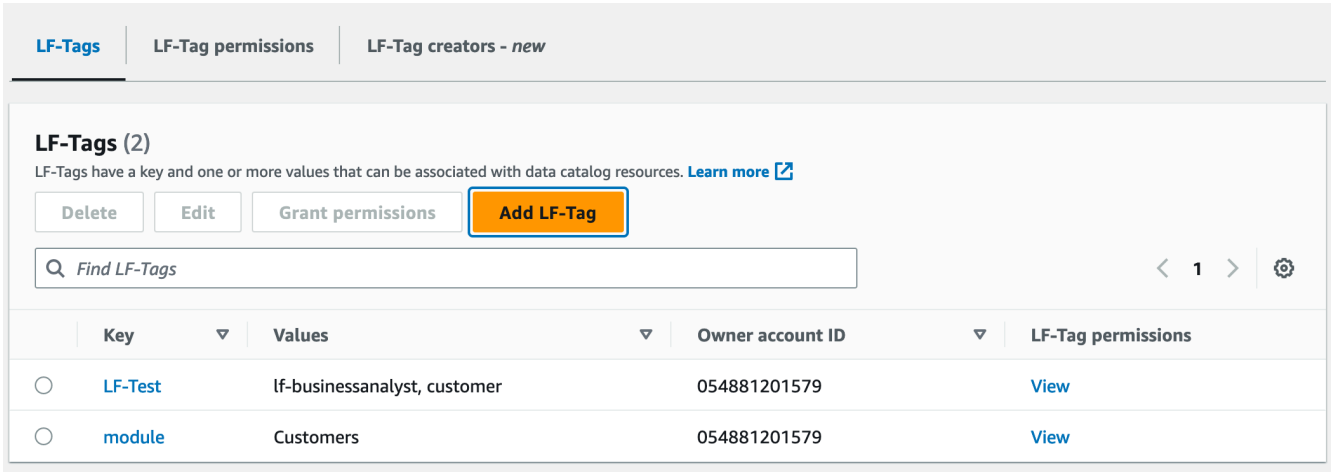
Console

Per elencare i tag LF (console)

1. Aprire la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.

   Accedi come creatore di LF-Tag, come amministratore del data lake o come responsabile a cui sono state concesse le autorizzazioni per i tag LF e che dispone dell'autorizzazione IAM. `lakeformation:ListLFTags`
2. Nel pannello di navigazione, sotto LF-tags e permessi, scegliete LF-Tags.

   Viene visualizzata la pagina LF-Tags.



Controlla la colonna Owner account ID per determinare i tag LF condivisi con il tuo account da un account esterno.

AWS CLI

Per elencare i tag LF () AWS CLI

- Esegui il comando seguente come amministratore del data lake o come responsabile a cui sono state concesse le autorizzazioni sui tag LF e che dispone dell'autorizzazione IAM. `lakeformation:ListLFTags`

```
aws lakeformation list-lf-tags
```

L'output è simile a quello riportato di seguito.

```
{
    "LFTags": [
        {
            "CatalogId": "111122223333",
            "TagKey": "level",
            "TagValues": [
                "director",
                "vp",
                "c-level"
            ]
        },
        {
            "CatalogId": "111122223333",
            "TagKey": "module",
            "TagValues": [
                "Orders",
                "Sales",
                "Customers"
            ]
        }
    ]
}
```

Per vedere anche i tag LF concessi da account esterni, includi l'opzione di comando. `--resource-share-type ALL`

```
aws lakeformation list-lf-tags --resource-share-type ALL
```

L'output è simile a quello riportato di seguito. Notate la `NextToken` chiave, che indica che ce ne sono altre da elencare.

```
{
    "LFTags": [
        {
            "CatalogId": "111122223333",
            "TagKey": "level",
            "TagValues": [
                "director",
                "vp",
                "c-level"
            ]
        },
        {
            "CatalogId": "111122223333",
            "TagKey": "module",
            "TagValues": [
                "Orders",
                "Sales",
                "Customers"
            ]
        }
    ],
    "NextToken": "eyJleHBpcmF0aW...ZXh0Ijp0cnVlfQ=="
}
```

Ripetete il comando e aggiungete l'`--next-token`argomento per visualizzare gli eventuali tag LF e tag LF locali rimanenti concessi da account esterni. I tag LF degli account esterni si trovano sempre su una pagina separata.

```
aws lakeformation list-lf-tags --resource-share-type ALL
--next-token eyJleHBpcmF0aW...ZXh0Ijp0cnVlfQ==
```

```
{
    "LFTags": [
        {
            "CatalogId": "123456789012",
```

```
            "TagKey": "region",
            "TagValues": [
                "central",
                "south"
            ]
        }
    ]
}
```

API

Puoi utilizzare gli SDK disponibili per Lake Formation per elencare i tag che il richiedente è autorizzato a visualizzare.

```
import boto3

client = boto3.client('lakeformation')
...

response = client.list_lf_tags(
    CatalogId='string',
    ResourceShareType='ALL',
    MaxResults=50'
)
```

Questo comando restituisce un `dict` oggetto con la seguente struttura:

```
{
    'LFTags': [
        {
            'CatalogId': 'string',
            'TagKey': 'string',
            'TagValues': [
                'string',
            ]
        },
    ],
    'NextToken': 'string'
}
```

Per ulteriori informazioni sulle autorizzazioni richieste, consulta Riferimento ai personaggi di Lake Formation e alle autorizzazioni IAM.

Assegnazione di tag LF alle risorse del Data Catalog

È possibile assegnare LF-tags alle risorse del Data Catalog (database, tabelle e colonne) per controllare l'accesso a tali risorse. Solo i principali a cui sono concessi i tag LF corrispondenti (e i principali a cui è concesso l'accesso con il metodo della risorsa denominata) possono accedere alle risorse.

Se una tabella eredita un tag LF da un database o una colonna eredita un tag LF da una tabella, è possibile sovrascrivere il valore ereditato assegnando un nuovo valore alla chiave LF-Tag.

Il numero massimo di LF-Tag che è possibile assegnare a una risorsa è 50.

Argomenti

- Requisiti per la gestione dei tag assegnati alle risorse
- Assegnate i tag LF a una colonna della tabella
- Assegnate i tag LF a una risorsa del catalogo dati
- Aggiornamento dei tag LF per una risorsa
- Rimozione del tag LF da una risorsa

Requisiti per la gestione dei tag assegnati alle risorse

Per assegnare un tag LF a una risorsa del catalogo dati, è necessario:

- Ottieni il `ASSOCIATE` permesso di Lake Formation sull'LF-Tag.
- Avere l'autorizzazione IAM`lakeformation:AddLFTagsToResource`.
- Have Glue: GetDatabase autorizzazione su un database Glue.
- Sii il proprietario della risorsa (creatore), disponi dell'autorizzazione `Super` Lake Formation sulla risorsa con l'`GRANT`opzione o disponi delle seguenti autorizzazioni con l'`GRANT`opzione:
  - Per i database nello stesso AWS account:`DESCRIBE`, `CREATE_TABLE``ALTER`, e `DROP`
  - Per i database in un account esterno:`DESCRIBE`, `CREATE_TABLE` e `ALTER`
  - Per tabelle (e colonne): `DESCRIBE``ALTER`,`DROP`,`INSERT`,`SELECT`, e `DELETE`

Inoltre, il tag LF e la risorsa a cui viene assegnato devono trovarsi nello stesso AWS account.

Per rimuovere un LF-tag da una risorsa del Data Catalog, è necessario soddisfare questi requisiti e disporre anche dell'autorizzazione IAM. `lakeformation:RemoveLFTagsFromResource`

Assegnate i tag LF a una colonna della tabella

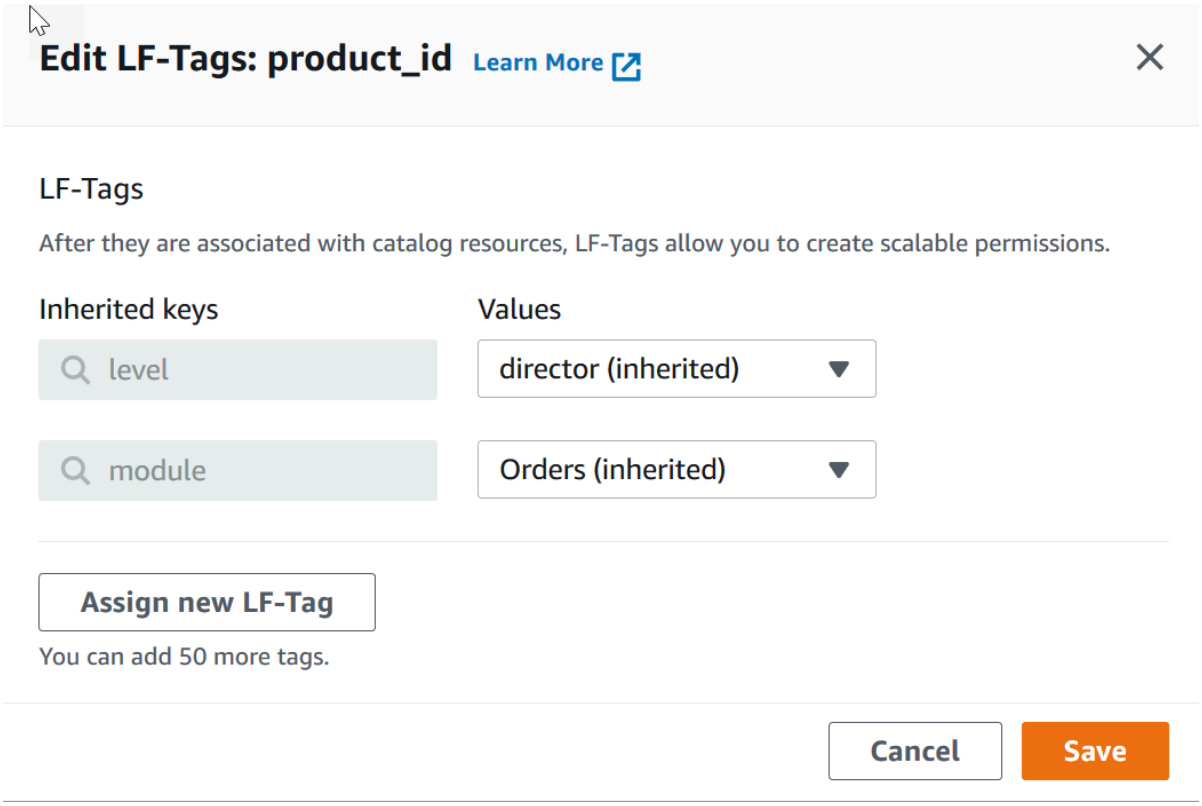
Per assegnare tag LF a una colonna della tabella (console)

1. Aprire la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.

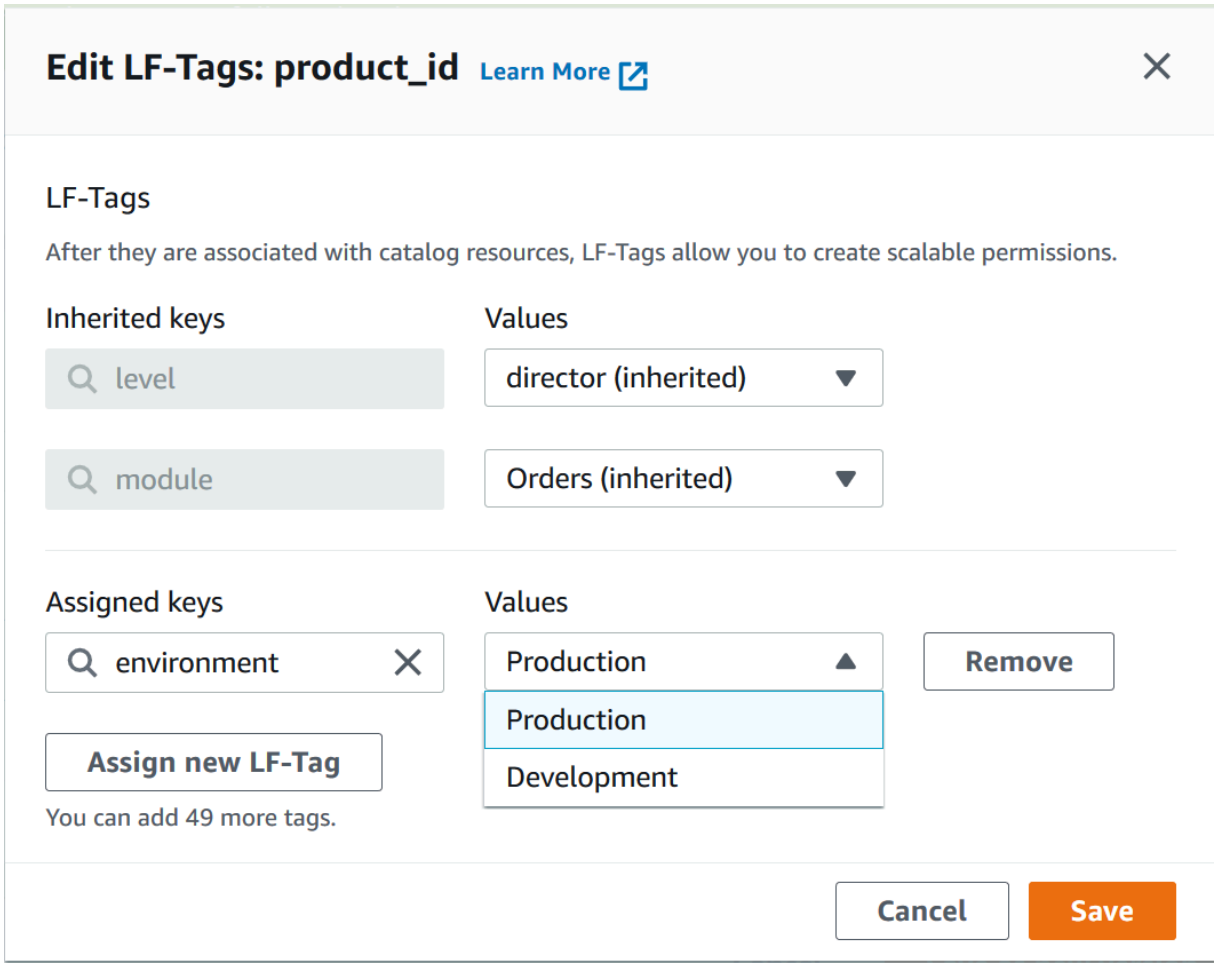
   Effettuate l'accesso come utente che soddisfa i requisiti sopra elencati.
2. Nel pannello di navigazione, seleziona Tabelle.
3. Scegli un nome per la tabella (non il pulsante di opzione accanto al nome della tabella).
4. Nella pagina dei dettagli della tabella, nella sezione Schema, scegli Modifica schema.
5. Nella pagina Modifica schema, seleziona una o più colonne, quindi scegli Modifica tag.

   > **Note**
   >
   > Se intendi aggiungere o eliminare colonne e salvare una nuova versione, esegui prima questa operazione. Quindi modificate i tag LF.

   Viene visualizzata la finestra di dialogo Modifica LF-Tag, che mostra tutti i tag LF ereditati dalla tabella.



6. (Facoltativo) Per l'elenco dei valori accanto a un campo Chiavi ereditate, scegliete un valore per sostituire il valore ereditato.
7. (Facoltativo) Scegliete Assegna nuovo tag LF. Quindi, per Tasti assegnati, scegliete una chiave e per Valori, scegliete un valore per la chiave.



8. (Facoltativo) Scegliete nuovamente Assegna nuovo tag LF per aggiungere un altro tag LF.
9. Selezionare Salva.

Assegnate i tag LF a una risorsa del catalogo dati

Console

Per assegnare LF-tags a un database o a una tabella del Data Catalog

1. Aprire la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.

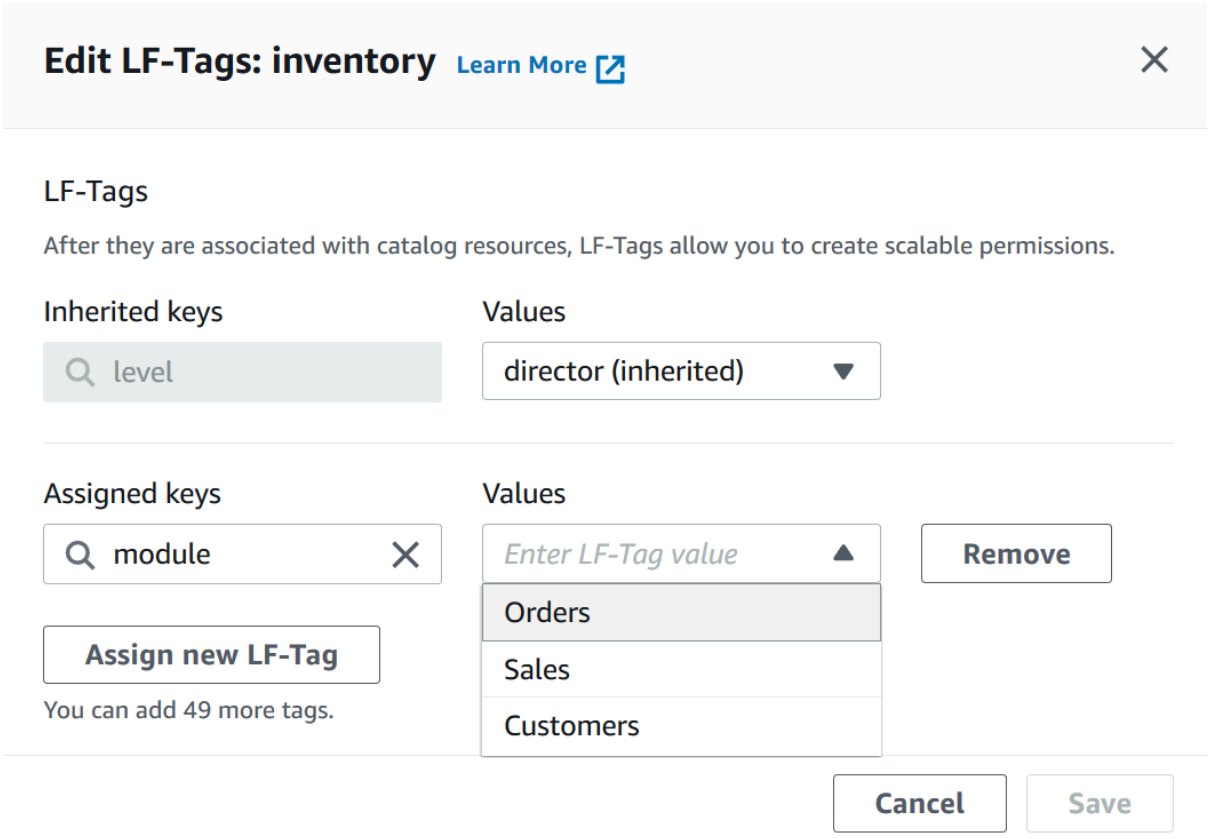
   Effettuate l'accesso come utente che soddisfa i requisiti elencati in precedenza.
2. Nel riquadro di navigazione, in Catalogo dati, esegui una delle seguenti operazioni:

   - Per assegnare tag LF ai database, scegliete Database.
   - Per assegnare i tag LF alle tabelle, scegliete Tabelle.

3. Scegliete un database o una tabella e nel menu Azioni scegliete Modifica tag.

   Viene visualizzata la finestra di dialogo Modifica tag LF: ***nome della risorsa***.

   Se una tabella eredita i tag LF dal database che la contiene, la finestra mostra i tag LF ereditati. Altrimenti, visualizza il testo «Non ci sono tag LF ereditati associati alla risorsa».



4. (Facoltativo) Se una tabella ha ereditato i tag LF, per l'elenco Valori accanto a un campo Chiavi ereditate, è possibile scegliere un valore per sostituire il valore ereditato.
5. Per assegnare nuovi LF-Tag, effettuate le seguenti operazioni:

   a. Scegliete Assegna nuovo LF-tag.

   b. Nel campo Chiavi assegnate, scegliete una chiave LF-Tag e nel campo Valori scegliete un valore.

   c. (Facoltativo) Scegliete nuovamente Assegna nuovo LF-tag per assegnare un tag LF aggiuntivo.
6. Selezionare Salva.

AWS CLI

Per assegnare tag LF a una risorsa del Data Catalog

- Esegui il comando `add-lf-tags-to-resource`.

  L'esempio seguente assegna il tag LF `module=orders` alla tabella del database. `orders erp` Per l'argomento viene utilizzata la sintassi della scorciatoia. `--lf-tags` La `CatalogID` proprietà for `--lf-tags` è facoltativa. Se non viene fornito, viene utilizzato l'ID di catalogo della risorsa (in questo caso, la tabella).

```
aws lakeformation add-lf-tags-to-resource --resource '{ "Table":
 {"DatabaseName":"erp", "Name":"orders"}}' --lf-tags
 CatalogId=111122223333,TagKey=module,TagValues=orders
```

  Quanto segue è l'output se il comando ha esito positivo.

```
{
    "Failures": []
}
```

  L'esempio successivo assegna due tag LF alla `sales` tabella e utilizza la sintassi JSON per l'argomento. `--lf-tags`

```
aws lakeformation add-lf-tags-to-resource --resource '{ "Table":
 {"DatabaseName":"erp", "Name":"sales"}}' --lf-tags '[{"TagKey":
 "module","TagValues": ["sales"]},{"TagKey": "environment","TagValues":
 ["development"]}]'
```

  L'esempio successivo assegna il tag LF alla colonna della tabella`level=director`. `total sales`

```
aws lakeformation add-lf-tags-to-resource --resource '{ "TableWithColumns":
 {"DatabaseName":"erp", "Name":"sales", "ColumnNames":["total"]}}' --lf-tags
 TagKey=level,TagValues=director
```

Aggiornamento dei tag LF per una risorsa

Per aggiornare un tag LF per una risorsa del catalogo dati () AWS CLI

- Utilizzate il `add-lf-tags-to-resource` comando, come descritto nella procedura precedente.

  L'aggiunta di un tag LF con la stessa chiave di un tag LF esistente, ma con un valore diverso, aggiorna il valore esistente.

Rimozione del tag LF da una risorsa

Per rimuovere un tag LF per una risorsa del Data Catalog () AWS CLI

- Esegui il comando `remove-lf-tags-from-resource`.

  Se una tabella ha un valore di tag LF che sostituisce il valore ereditato dal database principale, la rimozione di tale tag LF dalla tabella ripristina il valore ereditato. Questo comportamento si applica anche a una colonna che sostituisce i valori chiave ereditati dalla tabella.

  L'esempio seguente rimuove il tag LF `level=director` dalla colonna della `total` tabella. `sales` La `CatalogID` proprietà for `--lf-tags` è facoltativa. Se non viene fornito, viene utilizzato l'ID di catalogo della risorsa (in questo caso, la tabella).

```
aws lakeformation remove-lf-tags-from-resource
--resource ' { "TableWithColumns":
{ "DatabaseName": "erp",  "Name": "sales",  "ColumnNames":[ "total"]}}'
--lf-tags  CatalogId=111122223333,TagKey=level,TagValues=director
```

Visualizzazione dei tag LF assegnati a una risorsa

È possibile visualizzare i tag LF assegnati a una risorsa del Data Catalog. È necessario disporre dell'`ASSOCIATE`autorizzazione `DESCRIBE` o dell'autorizzazione su un LF-Tag per visualizzarlo.

Console

Per visualizzare i tag LF assegnati a una risorsa (console)

1. Aprire la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.

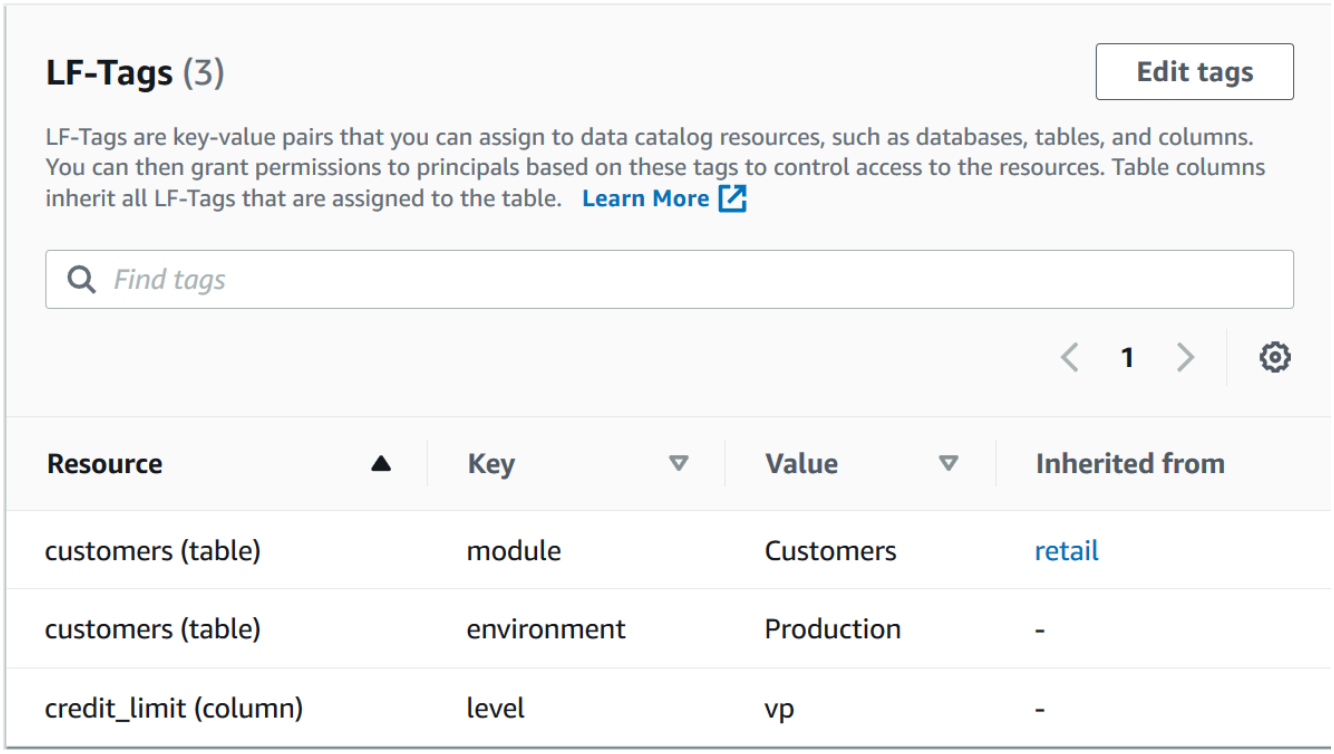
Accedi come amministratore del data lake, proprietario della risorsa o utente a cui sono state concesse le autorizzazioni di Lake Formation sulla risorsa.

2. Nel riquadro di navigazione, sotto l'intestazione Catalogo dati, esegui una delle seguenti operazioni:

   - Per visualizzare i tag LF assegnati a un database, scegliete Database.
   - Per visualizzare i tag LF assegnati a una tabella, scegliete Tabelle.

3. Nella pagina Tabelle o Database, scegliete il nome del database o della tabella. Quindi, nella pagina dei dettagli, scorri verso il basso fino alla sezione LF-Tags.

   La schermata seguente mostra i tag LF assegnati a una `customers` tabella, contenuta nel database. `retail` Il `module` tag LF viene ereditato dal database. Alla `credit_limit` colonna è assegnato il tag LF`level=vp`.



AWS CLI

Per visualizzare i tag LF assegnati a una risorsa () AWS CLI

- Utilizzare un comando simile al seguente:

```
aws lakeformation get-resource-lf-tags --show-assigned-lf-tags --
resource '{ "Table": {"CatalogId":"111122223333", "DatabaseName":"erp",
 "Name":"sales"}}'
```

Questo comando restituisce il seguente output.

```
{
    "TableTags": [
        {
            "CatalogId": "111122223333",
            "TagKey": "module",
            "TagValues": [
                "sales"
            ]
        },
        {
            "CatalogId": "111122223333",
            "TagKey": "environment",
            "TagValues": [
                "development"
            ]
        }
    ],
    "ColumnTags": [
        {
            "Name": "total",
            "Tags": [
                {
                    "CatalogId": "111122223333",
                    "TagKey": "level",
                    "TagValues": [
                        "director"
                    ]
                }
            ]
        }
    ]
}
```

Questo output mostra solo i tag LF assegnati in modo esplicito, non ereditati. Se volete vedere tutti i tag LF su tutte le colonne, compresi i tag LF ereditati, omettete l'opzione. `--show-assigned-lf-tags`

Visualizzazione delle risorse a cui è assegnato un LF-Tag

È possibile visualizzare tutte le risorse del Data Catalog a cui è assegnata una particolare chiave LF-Tag. A tale scopo, sono necessarie le seguenti autorizzazioni di Lake Formation:

- `Describe` o `Associate` sul tag LF.
- `Describe` o qualsiasi altra autorizzazione di Lake Formation sulla risorsa.

Inoltre, sono necessarie le seguenti autorizzazioni AWS Identity and Access Management (IAM):

- `lakeformation:SearchDatabasesByLFTags`
- `lakeformation:SearchTablesByLFTags`

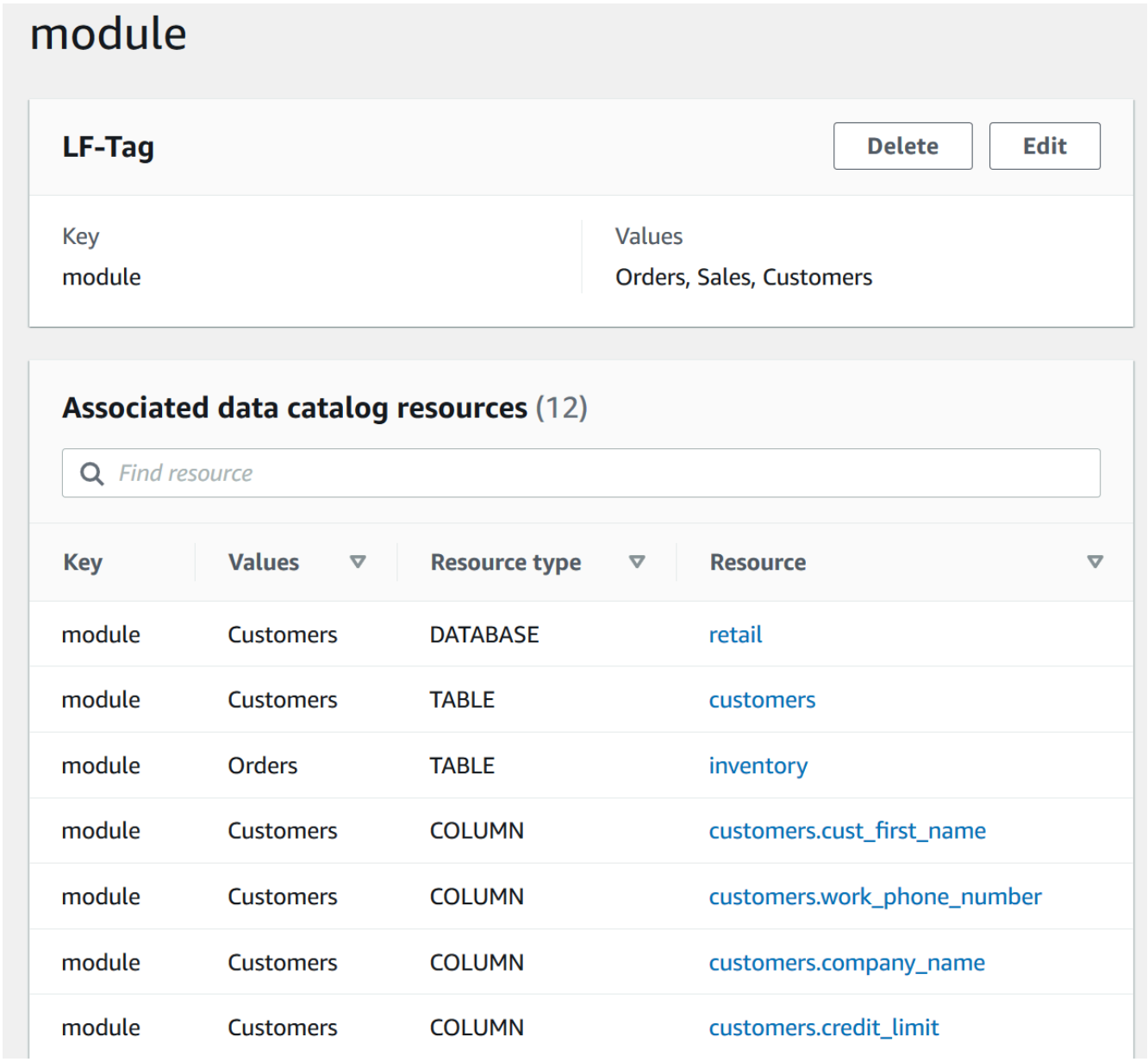
Console

Per visualizzare le risorse a cui è assegnato un tag LF (console)

1. Aprire la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.

   Accedi come amministratore del data lake o come utente che soddisfa i requisiti elencati in precedenza.
2. Nel riquadro di navigazione, sotto le rubriche Autorizzazioni e Ruoli e attività amministrativi, scegli LF-Tags.
3. Scegliete una chiave LF-Tag (non il pulsante di opzione accanto al nome della chiave).

   La pagina dei dettagli del tag LF mostra un elenco di risorse a cui è stato assegnato il tag LF.



AWS CLI

Per visualizzare le risorse a cui è assegnato un tag LF

- Esegui un comando `search-tables-by-lf-tags` or. `search-databases-by-lf-tags`

Example

L'esempio seguente elenca le tabelle e le colonne a cui è assegnato il `level=vp` tag LF. Per ogni tabella e colonna elencate, vengono emessi tutti i tag LF assegnati alla tabella o alla colonna, non solo l'espressione di ricerca.

```
aws lakeformation search-tables-by-lf-tags --expression
 TagKey=level,TagValues=vp
```

Per ulteriori informazioni sulle autorizzazioni richieste, consulta Riferimento ai personaggi di Lake Formation e alle autorizzazioni IAM.

Ciclo di vita di un tag LF

1. Il creatore del tag LF Michael crea un tag LF. `module=Customers`
2. Michael concede l'LF-Tag all'`Associate`ingegnere dei dati Eduardo. Concedere sovvenzioni implicitamente. `Associate` `Describe`
3. Michael concede `Super` `Custs` a Eduardo l'opzione grant, in modo che Eduardo possa assegnare dei tag LF al tavolo. Per ulteriori informazioni, consulta Assegnazione di tag LF alle risorse del Data Catalog.
4. Eduardo assegna il tag LF alla tabella. `module=customers` `Custs`
5. Michael concede la seguente concessione all'ingegnere dei dati Sandra (in pseudo-codice).

```
GRANT (SELECT, INSERT ON TABLES) ON TAGS module=customers TO Sandra WITH GRANT OPTION
```

6. Sandra concede la seguente concessione all'analista di dati Maria.

```
GRANT (SELECT ON TABLES) ON TAGS module=customers TO Maria
```

Maria ora può eseguire interrogazioni sul tavolo. `Custs`

Consulta anche

- Controllo dell'accesso ai metadati

Confronto tra il controllo degli accessi basato su tag Lake Formation e il controllo degli accessi basato sugli attributi IAM

Il controllo dell'accesso basato su attributi (ABAC) è una strategia di autorizzazione che definisce le autorizzazioni in base agli attributi. In AWS, tali attributi sono denominati tag. È possibile collegare dei tag alle risorse IAM, tra cui le entità IAM (utenti o ruoli), e alle risorse AWS. È possibile creare una singola policy ABAC o un piccolo insieme di policy per i principali IAM. Queste policy ABAC possono essere definite affinché autorizzino le operazioni quando il tag dell'entità corrisponde al tag della risorsa. La strategia ABAC è utile in ambienti soggetti a una rapida crescita e aiuta in situazioni in cui la gestione delle policy diventa impegnativa.

I team di sicurezza e governance del cloud utilizzano IAM per definire le policy di accesso e le autorizzazioni di sicurezza per tutte le risorse, inclusi i bucket Amazon S3, le istanze Amazon EC2 e tutte le risorse a cui puoi fare riferimento con un ARN. Le policy IAM definiscono autorizzazioni ampie (granulari) per le risorse del data lake, ad esempio per consentire o negare l'accesso a livello di bucket o prefisso Amazon S3 o a livello di database. Per ulteriori informazioni su IAM ABAC, consulta A cosa serve ABAC? AWS nella Guida per l'utente di IAM.

Ad esempio, è possibile creare tre ruoli associando a essi un tag con la chiave `project-access` e impostando il valore del tag del primo ruolo a `Dev`, del secondo a `Marketing` e del terzo a `Support`. Assegna tag con il valore appropriato alle risorse. È quindi possibile utilizzare una singola policy che consenta l'accesso quando il ruolo e la risorsa sono contrassegnati con lo stesso valore del tag `project-access`.

I team di governance dei dati utilizzano Lake Formation per definire autorizzazioni granulari per specifiche risorse di data lake. I tag LF sono assegnati alle risorse del Data Catalog (database, tabelle e colonne) e sono concessi ai principali. Un principale con tag LF che corrispondono ai tag LF di una risorsa può accedere a quella risorsa. Le autorizzazioni Lake Formation sono secondarie rispetto alle autorizzazioni IAM. Ad esempio, se le autorizzazioni IAM non consentono a un utente di accedere a un data lake, Lake Formation non concede l'accesso a nessuna risorsa all'interno di quel data lake a quell'utente, anche se il principale e la risorsa hanno tag LF corrispondenti.

Il controllo degli accessi basato su tag di Lake Formation (LF-TBAC) funziona con IAM ABAC per fornire livelli aggiuntivi di autorizzazioni per i dati e le risorse di Lake Formation.

- Le autorizzazioni TBAC di Lake Formation si adattano all'innovazione. Non è più necessario che un amministratore aggiorni le policy esistenti per consentire l'accesso a nuove risorse. Ad esempio, supponiamo di utilizzare una strategia IAM ABAC con il `project-access` tag per fornire l'accesso a database specifici all'interno di Lake Formation. Utilizzando LF-TBAC, il tag

LF `Project=SuperApp` viene assegnato a tabelle o colonne specifiche e lo stesso tag LF viene concesso a uno sviluppatore per quel progetto. Tramite IAM, lo sviluppatore può accedere al database e le autorizzazioni LF-TBAC garantiscono allo sviluppatore un ulteriore accesso a tabelle o colonne specifiche all'interno delle tabelle. Se viene aggiunta una nuova tabella al progetto, l'amministratore di Lake Formation deve solo assegnare il tag alla nuova tabella affinché lo sviluppatore possa accedere alla tabella.

- Lake Formation TBAC richiede meno policy IAM. Poiché utilizzi le policy IAM per garantire un accesso di alto livello alle risorse di Lake Formation e Lake Formation TBAC per gestire un accesso più preciso ai dati, crei meno policy IAM.
- Utilizzando Lake Formation TBAC, i team possono cambiare e crescere rapidamente. Questo perché le autorizzazioni per le nuove risorse vengono concesse automaticamente in base agli attributi. Ad esempio, se un nuovo sviluppatore si unisce al progetto, è facile concedere a tale sviluppatore l'accesso associando il ruolo IAM all'utente e quindi assegnandogli i tag LF richiesti. Non è necessario modificare la policy IAM per supportare un nuovo progetto o creare nuovi LF-tag.
- Autorizzazioni più dettagliate sono possibili utilizzando Lake Formation TBAC. Le policy IAM garantiscono l'accesso alle risorse di primo livello, come i database o le tabelle del Data Catalog. Utilizzando Lake Formation TBAC, puoi concedere l'accesso a tabelle o colonne specifiche che contengono valori di dati specifici.

> **Note**
>
> I tag IAM non sono gli stessi dei tag LF. Questi tag non sono intercambiabili. I tag LF vengono utilizzati per concedere i permessi di Lake Formation e i tag IAM vengono utilizzati per definire le politiche IAM.

Concessione, revoca ed elenco delle autorizzazioni per i valori LF-Tag

È possibile concedere le `Drop` `Alter` autorizzazioni sui tag LF ai responsabili per gestire le espressioni di valore dei tag LF. È inoltre possibile concedere `Describe` `Grant with LF-Tag expressions` le autorizzazioni relative ai tag LF ai principali per visualizzare i tag LF e assegnarli alle risorse del Data Catalog (database, tabelle e colonne). `Associate` Quando i tag LF vengono assegnati alle risorse del Data Catalog, è possibile utilizzare il metodo di controllo degli accessi basato su tag Lake Formation (LF-TBAC) per proteggere tali risorse. Per ulteriori informazioni, consulta Controllo degli accessi basato su tag Lake Formation.

È possibile concedere queste autorizzazioni con l'opzione di concessione in modo che altri responsabili possano concederle. Le `Associate` autorizzazioni `Grant with LF-Tag expressions``Describe`, e sono spiegate in. Aggiungi creatori di tag LF

È possibile concedere le `Associate` autorizzazioni `Describe` and su un tag LF a un account esterno. AWS Un amministratore di data lake in quell'account può quindi concedere tali autorizzazioni ad altri responsabili dell'account. I responsabili a cui l'amministratore del data lake dell'account esterno concede l'`Associate`autorizzazione possono quindi assegnare LF-Tags alle risorse del Data Catalog che hai condiviso con il loro account.

Quando si concede a un account esterno, è necessario includere l'opzione di concessione.

Puoi concedere le autorizzazioni sui tag LF utilizzando la console Lake Formation, l'API o il AWS Command Line Interface ().AWS CLI

Argomenti

- Elencare le autorizzazioni LF-Tag utilizzando la console
- Concessione delle autorizzazioni LF-Tag tramite la console
- Concessione, revoca ed elenco delle autorizzazioni LF-Tag utilizzando il AWS CLI

Per ulteriori informazioni, consulta Gestione dei tag LF per il controllo dell'accesso ai metadati e Controllo degli accessi basato su tag Lake Formation.

Elencare le autorizzazioni LF-Tag utilizzando la console

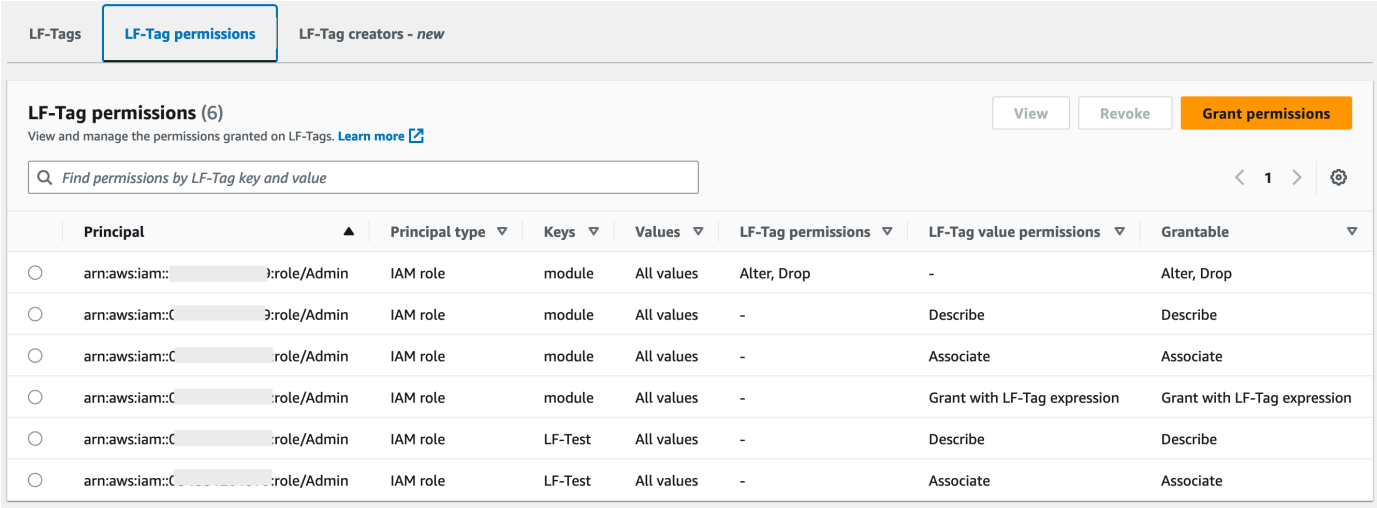
Puoi usare la console Lake Formation per visualizzare le autorizzazioni concesse sui tag LF. Devi essere un creatore di LF-Tag, un amministratore di data lake o avere l'`Associate`autorizzazione `Describe` o l'autorizzazione su un LF-Tag per vederlo.

Per elencare le autorizzazioni LF-Tag (console)

1. Aprire la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.

   Accedi come creatore di LF-Tag, amministratore di data lake o come utente a cui sono state concesse le`Drop`, `AlterAssociate`, o le `Describe` autorizzazioni per i tag LF.
2. Nel pannello di navigazione, in Autorizzazioni, scegli LF-Tag e permessi e scegli la sezione Autorizzazioni LF-Tag.

La sezione delle autorizzazioni LF-Tag mostra una tabella che contiene il principale, le chiavi dei tag, i valori e le autorizzazioni.



Concessione delle autorizzazioni LF-Tag tramite la console

I passaggi seguenti spiegano come concedere le autorizzazioni sui tag LF utilizzando la pagina Grant LF-tag permissions sulla console Lake Formation. La pagina è suddivisa nelle seguenti sezioni:

- Tipi di autorizzazione: il tipo di autorizzazione da concedere.
- Principali: gli utenti, i ruoli o gli AWS account a cui concedere le autorizzazioni.
- LF-Tags — I tag LF su cui concedere le autorizzazioni.
- Autorizzazioni: le autorizzazioni da concedere.

Apri la pagina Concedi le autorizzazioni del tag LF

1. Aprire la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.

   Accedi come creatore di LF-Tag, amministratore di data lake o come utente Le autorizzazioni LF-Tag o la coppia chiave-valore LF-Tag su LF-Tag sono state concesse con l'opzione. `Grant`
2. Nel pannello di navigazione, scegli LF-Tag e permessi, scegli la sezione Autorizzazioni LF-Tag.
3. Scegli Concedi autorizzazioni.

Specificate il tipo di autorizzazione

Nella sezione Tipo di autorizzazione, scegli un tipo di autorizzazione.

Autorizzazioni LF-Tag

Scegli le autorizzazioni LF-Tag per consentire ai mandanti di aggiornare i valori dei tag LF o eliminare i tag LF.

Autorizzazioni per la coppia chiave-valore LF-Tag

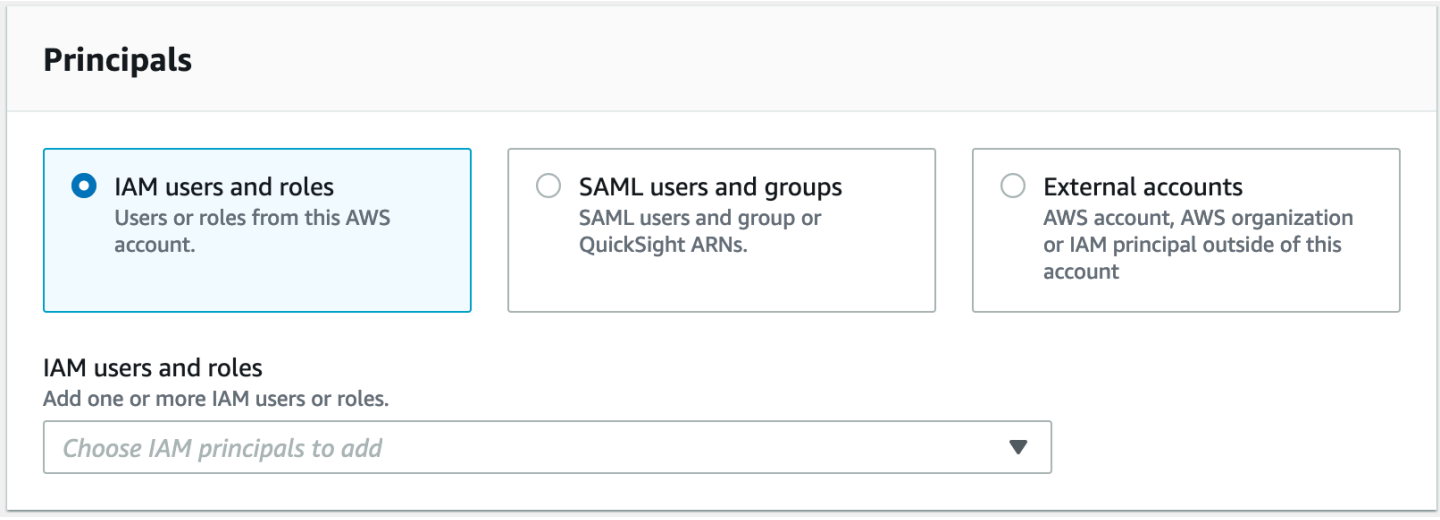
Scegliete i permessi della coppia chiave-valore LF-Tag per consentire ai mandanti di assegnare LF-Tags alle risorse del Data Catalog, visualizzare i tag e i valori LF-Tag e concedere ai principali autorizzazioni basate sui tag LF sulle risorse del Data Catalog.

Le opzioni disponibili nelle seguenti sezioni dipendono dal tipo di autorizzazione.

Specificare i principali

> **Note**
>
> Non puoi concedere autorizzazioni LF-Tag (`AlterandDrop`) ad account o responsabili esterni di un altro account.

Nella sezione Principali, scegli un tipo principale e specifica i principali a cui concedere le autorizzazioni.



Utenti e ruoli IAM

Scegli uno o più utenti o ruoli dall'elenco degli utenti e dei ruoli IAM.

Utenti e gruppi SAML

Per QuickSight utenti e gruppi SAML e Amazon, inserisci uno o più Amazon Resource Names (ARN) per utenti o gruppi federati tramite SAML o ARN per utenti o gruppi Amazon. QuickSight Premi Invio dopo ogni ARN.

Per informazioni su come costruire gli ARN, vedere. Comandi di concessione e AWS CLI revoca di Lake Formation

> **Note**
>
> L'integrazione di Lake Formation con Amazon QuickSight è supportata solo per Amazon QuickSight Enterprise Edition.

Account esterni

Per l'AWS account, inserisci uno o più ID AWS account validi. Premi Invio dopo ogni ID.

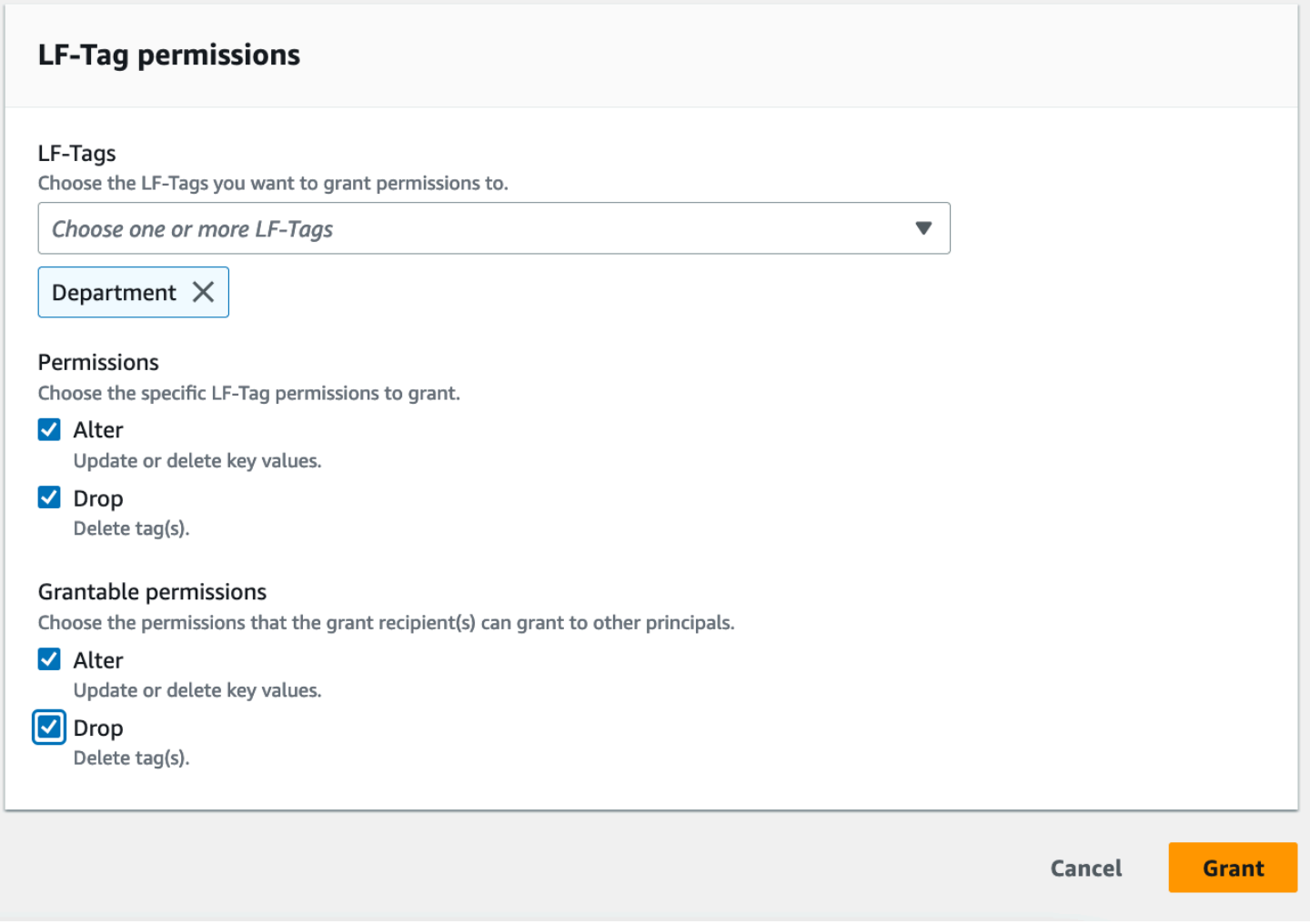
Un ID dell'organizzazione è composto da «o-» seguito da 10-32 lettere o cifre minuscole.

L'ID di un'unità organizzativa inizia con «ou-» seguito da 4 a 32 lettere o cifre minuscole (l'ID della radice che contiene l'unità organizzativa). Questa stringa è seguita da un secondo trattino «-» e da 8 a 32 lettere o cifre minuscole aggiuntive.

Per il principale IAM, inserisci l'ARN per l'utente o il ruolo IAM.

Specificate i tag LF

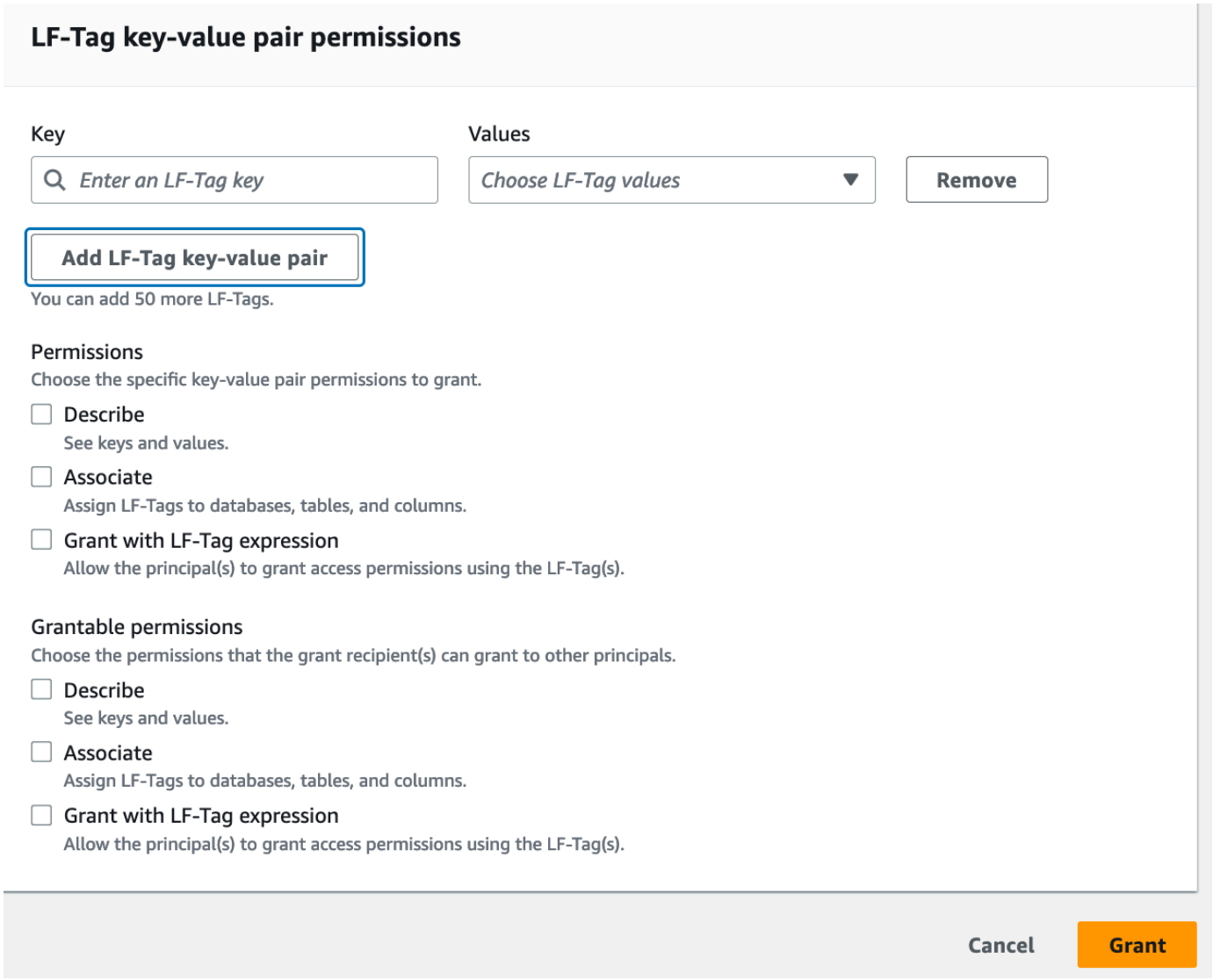
Per concedere le autorizzazioni sui tag LF, nella sezione Autorizzazioni dei tag LF, specifica i tag LF su cui concedere le autorizzazioni.



- Scegli uno o più tag LF utilizzando il menu a discesa.

Specificate le coppie chiave-valore del tag LF

1. Per concedere i permessi sulle coppie chiave-valore LF-Tag, (devi prima scegliere Scegli i permessi della coppia chiave-valore LF-Tag come Tipo di autorizzazione) scegli Aggiungi coppia chiave-valore LF-Tag per visualizzare la prima riga di campi per specificare la chiave e i valori del tag LF.



2. Posizionate il cursore nel campo Chiave, facoltativamente iniziate a digitare per restringere l'elenco di selezione e selezionate una chiave LF-Tag.
3. Nell'elenco Valori, selezionate uno o più valori, quindi premete Tab oppure fate clic o toccate all'esterno del campo per salvare i valori selezionati.

> **Note**
>
> Se una delle righe dell'elenco Valori è attiva, premendo Invio si seleziona o deseleziona la casella di controllo.

I valori selezionati vengono visualizzati come riquadri sotto l'elenco Valori. Scegli ✖ per rimuovere un valore. Scegli Rimuovi per rimuovere l'intero tag LF.

4. Per aggiungere un altro tag LF, scegliete nuovamente Aggiungi tag LF e ripetete i due passaggi precedenti.

Specificate le autorizzazioni

Questa sezione mostra le autorizzazioni del tag LF o le autorizzazioni con valore del tag LF in base al tipo di autorizzazione scelto nel passaggio precedente.

A seconda del tipo di autorizzazione che hai scelto di concedere, seleziona le autorizzazioni LF-Tag o la coppia chiave-valore LF-Tag, le autorizzazioni e le autorizzazioni concedibili.

1. In Autorizzazioni LF-Tag, seleziona le autorizzazioni da concedere.

   La concessione di Drop and Alter implica la concessione implicita di Describe.

   È necessario concedere le autorizzazioni Alter and Drop su tutti i valori dei tag.
2. In LT-Tag key-value permissions, seleziona le autorizzazioni da concedere.

   Granting Associate concede implicitamente Describe. Scegliete Grant with LF-Tag expression per consentire al destinatario della concessione di concedere o revocare le autorizzazioni di accesso alle risorse del Data Catalog utilizzando il metodo LF-TBAC.
3. (Facoltativo) In Autorizzazioni concedibili, seleziona le autorizzazioni che il destinatario della sovvenzione può concedere ad altri responsabili del proprio account. AWS
4. Scegli Concessione.

Concessione, revoca ed elenco delle autorizzazioni LF-Tag utilizzando il AWS CLI

È possibile concedere, revocare ed elencare i permessi sui tag LF utilizzando (). AWS Command Line Interface AWS CLI

Per elencare le autorizzazioni LF-Tag ()AWS CLI

- Immettete un comando. `list-permissions` Devi essere il creatore del tag LF, un amministratore del data lake o avere l'`Grant with LF-Tag permissions`autorizzazione`Drop`,, `Alter` `Describe``Associate`, per un tag LF per vederlo.

  Il comando seguente richiede tutti i tag LF per i quali disponete delle autorizzazioni.

  ```
  aws lakeformation list-permissions --resource-type LF_TAG
  ```

Di seguito è riportato un esempio di output per un amministratore di data lake, che vede tutti i tag LF concessi a tutti i principali. Gli utenti non amministrativi vedono solo i tag LF concessi loro. Le autorizzazioni LF-Tag concesse da un account esterno vengono visualizzate in una pagina dei risultati separata. Per vederli, ripetete il comando e inserite nell'`--next-token`argomento il token restituito dalla precedente esecuzione del 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"
}
```

È possibile elencare tutte le concessioni per una chiave LF-Tag specifica. Il comando seguente restituisce tutte le autorizzazioni concesse per il tag LF. `module`

```
aws lakeformation list-permissions --resource-type LF_TAG --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

È inoltre possibile elencare i valori del tag LF concessi a un principale specifico per un tag LF specifico. Quando si fornisce l'`--principal`argomento, è necessario fornire l'argomento. `--resource` Pertanto, il comando può richiedere efficacemente solo i valori concessi a un principale specifico per una chiave LF-Tag specifica. Il comando seguente mostra come eseguire questa operazione per la chiave principale `datalake_user1` e per la chiave LF-Tag. `module`

```
aws lakeformation list-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --resource-type LF_TAG --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

Di seguito è riportato un output di esempio.

```
{
    "PrincipalResourcePermissions": [
        {
            "Principal": {
                "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"
            },
            "Resource": {
                "LFTag": {
```

```
                    "CatalogId": "111122223333",
                    "TagKey": "module",
                    "TagValues": [
                        "Orders",
                        "Sales"
                    ]
                }
            },
            "Permissions": [
                "ASSOCIATE"
            ],
            "PermissionsWithGrantOption": []
        }
    ]
}
```

Per concedere i permessi sui tag LF ()AWS CLI

1. Utilizzare un comando simile al seguente: Questo esempio concede all'utente l'`Associate`autorizzazione per l'uso `datalake_user1` del tag LF con la chiave. `module` Concede le autorizzazioni per visualizzare e assegnare tutti i valori per quella chiave, come indicato dall'asterisco (*).

   ```
   aws lakeformation grant-permissions --principal
    DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
   datalake_user1 --permissions "ASSOCIATE" --resource '{ "LFTag":
    {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
   ```

   La concessione dell'autorizzazione implica la concessione implicita dell'`Associate`autorizzazione. `Describe`

   L'esempio successivo concede `Associate` all' AWS account esterno 1234-5678-9012 sul tag LF con la chiave, con l'opzione grant. `module` Concede le autorizzazioni per visualizzare e assegnare solo i valori e. `sales` `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. La concessione dell'autorizzazione implica la concessione implicita `GrantWithLFTagExpression` dell'autorizzazione. `Describe`

   L'esempio successivo concede `GrantWithLFTagExpression` a un utente il tag LF con la chiave`module`, con l'opzione grant. Concede le autorizzazioni per visualizzare e concedere le autorizzazioni sulle risorse del Data Catalog utilizzando solo i valori e. `sales` `orders`

   ```
   aws lakeformation grant-permissions --principal
    DataLakePrincipalIdentifier=111122223333 --permissions "GrantWithLFTagExpression"
    --permissions-with-grant-option "GrantWithLFTagExpression" --resource '{ "LFTag":
    {"CatalogId":"111122223333","TagKey":"module","TagValues":["sales", "orders"]}}'
   ```

3. L'esempio successivo concede `Drop` le autorizzazioni a un utente sul tag LF con la chiave, con l'opzione grant. `module` Concede le autorizzazioni per eliminare il tag LF. Per eliminare un tag LF, sono necessarie le autorizzazioni su tutti i valori di quella chiave.

   ```
   aws lakeformation grant-permissions --principal
    DataLakePrincipalIdentifier=111122223333 --permissions "DROP"
    --permissions-with-grant-option "DROP" --resource '{ "LFTag":
    {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
   ```

4. L'esempio successivo concede `Alter` le autorizzazioni all'utente sul tag LF con la chiave, con l'opzione grant. `module` Concede le autorizzazioni per eliminare il tag LF. Per aggiornare un LF-tag, sono necessarie le autorizzazioni su tutti i valori di quella chiave.

   ```
   aws lakeformation grant-permissions --principal
    DataLakePrincipalIdentifier=111122223333 --permissions "ALTER"
    --permissions-with-grant-option "ALTER" --resource '{ "LFTag":
    {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
   ```

Per revocare le autorizzazioni su LF-Tags ()AWS CLI

- Utilizzare un comando simile al seguente: Questo esempio revoca l'`Associate`autorizzazione sul tag LF con la chiave dell'utente. `module` `datalake_user1`

  ```
  aws lakeformation revoke-permissions --principal
   DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
  datalake_user1 --permissions "ASSOCIATE" --resource '{ "LFTag":
   {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
  ```

Concessione delle autorizzazioni per il data lake utilizzando il metodo LF-TBAC

Puoi concedere le autorizzazioni `DESCRIBE` e `ASSOCIATE` Lake Formation sui tag LF ai mandanti in modo che possano visualizzare i tag LF e assegnarli alle risorse del Data Catalog (database, tabelle, viste e colonne). Quando i tag LF vengono assegnati alle risorse del Data Catalog, è possibile utilizzare il metodo di controllo degli accessi basato su tag Lake Formation (LF-TBAC) per proteggere tali risorse. Per ulteriori informazioni, consulta Controllo degli accessi basato su tag Lake Formation.

Inizialmente, solo l'amministratore del data lake può concedere queste autorizzazioni. Se l'amministratore del data lake concede queste autorizzazioni con l'opzione di concessione, altri responsabili possono concederle. Le `ASSOCIATE` autorizzazioni `DESCRIBE` e sono spiegate in. Buone pratiche e considerazioni per il controllo degli accessi basato su tag Lake Formation

È possibile concedere le `ASSOCIATE` autorizzazioni `DESCRIBE` and su un tag LF a un account esterno. AWS Un amministratore di data lake in quell'account può quindi concedere tali autorizzazioni ad altri responsabili dell'account. I responsabili a cui l'amministratore del data lake dell'account esterno concede l'`ASSOCIATE`autorizzazione possono quindi assegnare LF-Tags alle risorse del Data Catalog che hai condiviso con il loro account.

Quando si concede a un account esterno, è necessario includere l'opzione di concessione.

È possibile concedere le autorizzazioni sui tag LF utilizzando la AWS Lake Formation console, l'API o (). AWS Command Line Interface AWS CLI

Argomenti

- Concessione delle autorizzazioni di Data Catalog

Consulta anche

- Concessione, revoca ed elenco delle autorizzazioni per i valori LF-Tag
- Gestione dei tag LF per il controllo dell'accesso ai metadati
- Controllo degli accessi basato su tag Lake Formation

Concessione delle autorizzazioni di Data Catalog

Usa la console Lake Formation o AWS CLI concedi le autorizzazioni di Lake Formation su database, tabelle, viste e colonne di Data Catalog utilizzando il metodo di controllo degli accessi basato su tag Lake Formation (LF-TBAC).

Console

I passaggi seguenti spiegano come concedere le autorizzazioni utilizzando il metodo di controllo degli accessi basato su tag Lake Formation (LF-TBAC) e la pagina Grant data lake permissions sulla console Lake Formation. La pagina è suddivisa nelle seguenti sezioni:

- Principi: utenti, ruoli e autorizzazioni Account AWS a cui concedere le autorizzazioni.
- Tag LF o risorse del catalogo: database, tabelle o collegamenti a risorse a cui concedere le autorizzazioni.
- Autorizzazioni: autorizzazioni da concedere a The Lake Formation.

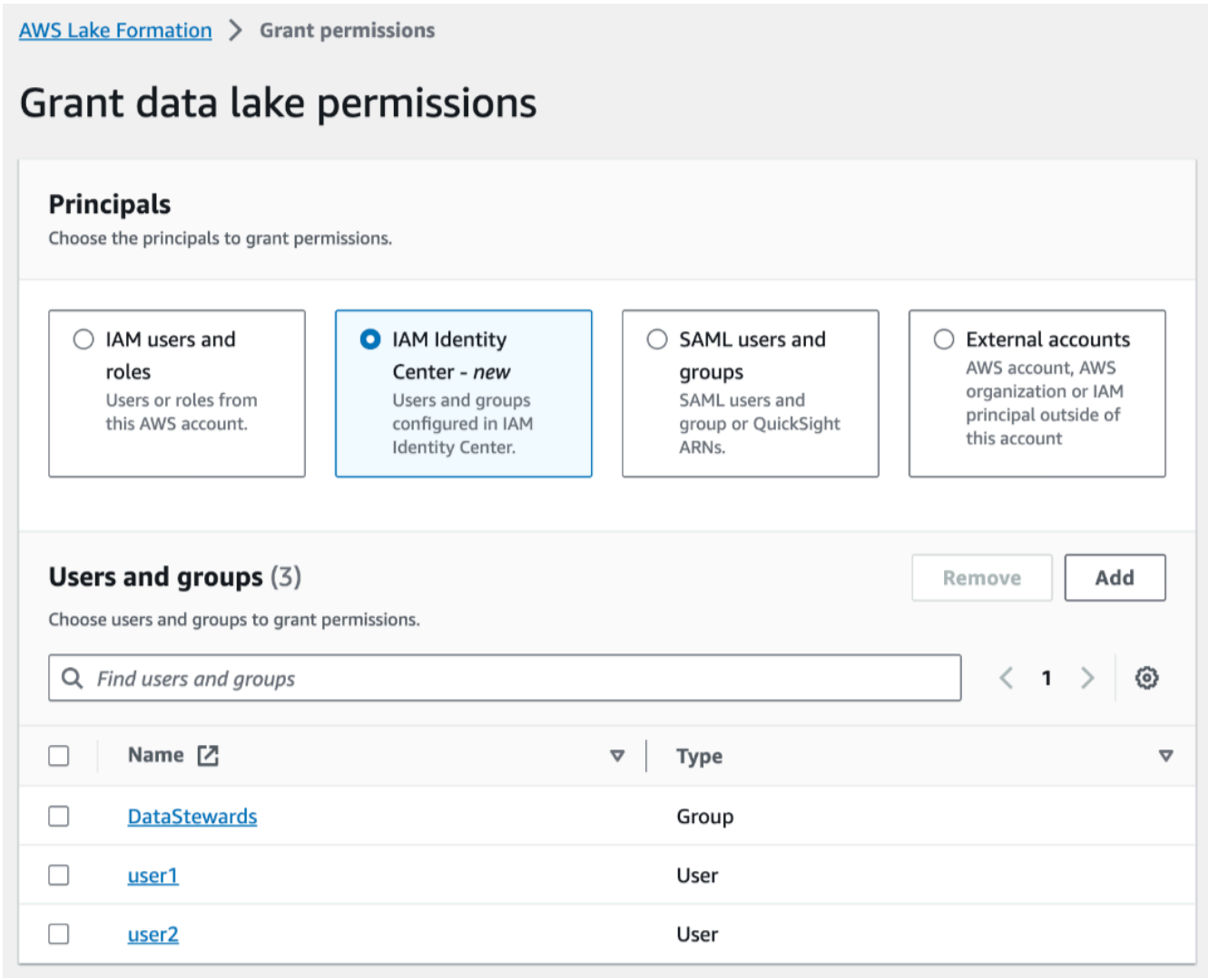
1. Apri la pagina Concedi le autorizzazioni del data lake.

   Apri la AWS Lake Formation console all'indirizzo https://console.aws.amazon.com/lakeformation/ e accedi come amministratore del data lake o come utente a cui sono state concesse le autorizzazioni Lake Formation sulle risorse del Data Catalog tramite LF-TBAC con l'opzione di concessione.

   Nel pannello di navigazione, in Autorizzazioni, scegli Autorizzazioni Data Lake. Quindi scegli Concedi.

2. Specificate i principali.

   Nella sezione Principali, scegli un tipo principale, quindi specifica i principali a cui concedere le autorizzazioni.



Utenti e ruoli IAM

Scegli uno o più utenti o ruoli dall'elenco degli utenti e dei ruoli IAM.

IAM Identity Center

Scegli uno o più utenti o dall'elenco Utenti e gruppi.

Utenti e gruppi SAML

Per QuickSight utenti e gruppi SAML e Amazon, inserisci uno o più Amazon Resource Names (ARN) per utenti o gruppi federati tramite SAML o ARN per utenti o gruppi Amazon. QuickSight Premi Invio dopo ogni ARN.

Per informazioni su come costruire gli ARN, vedere. Comandi di concessione e AWS CLI revoca di Lake Formation

> **Note**
>
> L'integrazione di Lake Formation con Amazon QuickSight è supportata solo per Amazon QuickSight Enterprise Edition.

Account esterni

Per Account AWS, AWS organizzazione o responsabile IAM inserisci uno o più Account AWS ID, ID organizzazione, ID unità organizzative o ARN validi per l'utente o il ruolo IAM. Premi Invio dopo ogni ID.

Un ID dell'organizzazione è composto da «o-» seguito da 10-32 lettere o cifre minuscole.

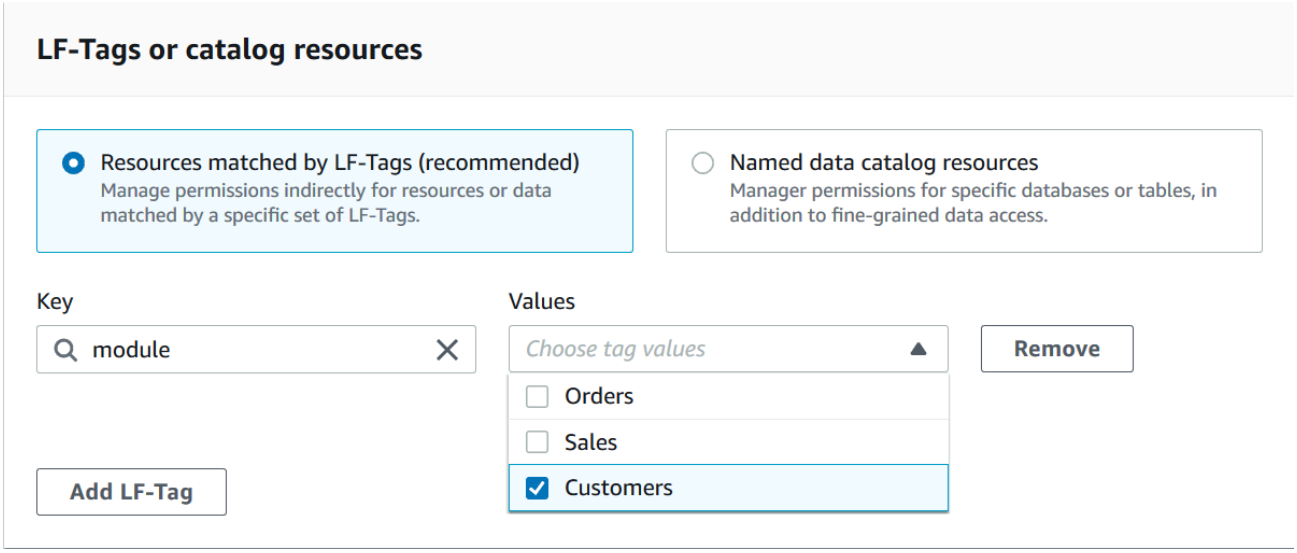
L'ID di un'unità organizzativa inizia con «ou-» seguito da 4 a 32 lettere o cifre minuscole (l'ID della radice che contiene l'unità organizzativa). Questa stringa è seguita da un secondo trattino «-» e da 8 a 32 lettere o cifre minuscole aggiuntive.

3. Specificate i tag LF.

Assicuratevi che sia selezionata l'opzione Risorse abbinate ai tag LF. Scegli Aggiungi tag LF.

1. Scegliete una chiave e dei valori LF-Tag.

Se scegliete più di un valore, state creando un'espressione LF-Tag con un operatore. `OR` Ciò significa che se uno qualsiasi dei valori del tag LF corrisponde a un tag LF assegnato a una risorsa del Data Catalog, all'utente vengono concesse le autorizzazioni sulla risorsa.



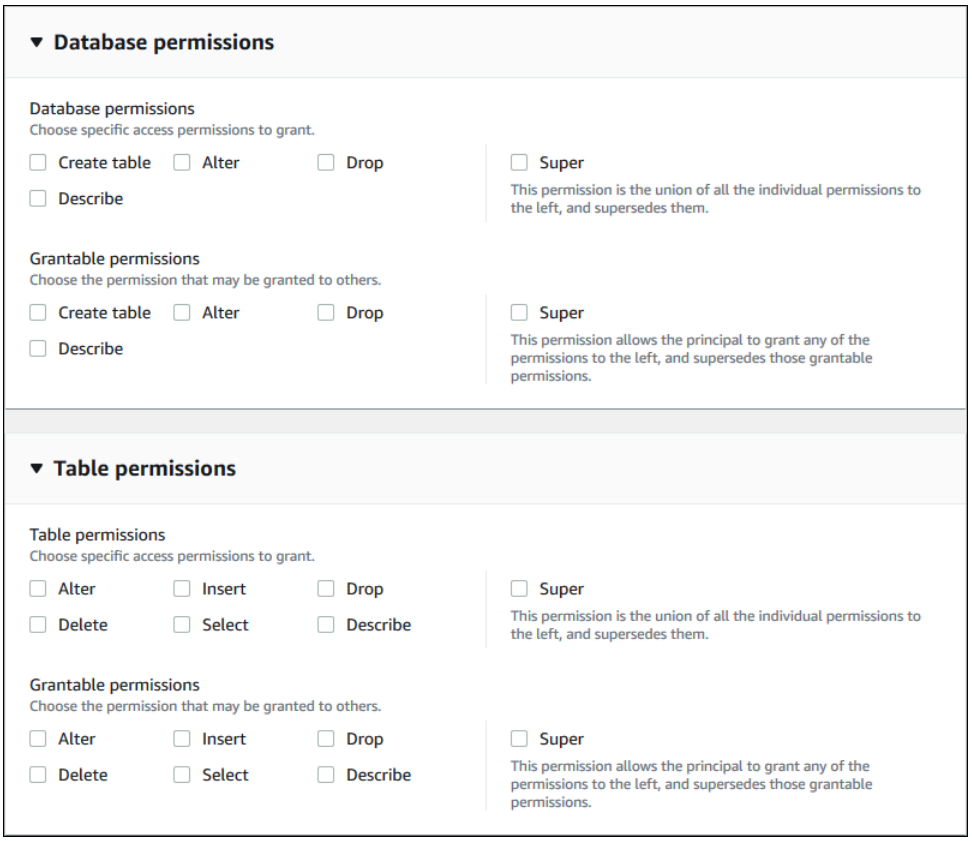
2. (Facoltativo) Scegliete nuovamente Aggiungi tag LF per specificare un altro tag LF.

   Se specificate più di un tag LF, state creando un'espressione di tag LF con un operatore. `AND` Al principale vengono concesse le autorizzazioni su una risorsa del Data Catalog solo se alla risorsa è stato assegnato un tag LF corrispondente per ogni tag LF nell'espressione del tag LF.

4. Specificare le autorizzazioni.

   Specificate le autorizzazioni che desiderate concedere al principale per le risorse del Data Catalog corrispondenti. Le risorse corrispondenti sono quelle a cui sono stati assegnati tag LF che corrispondono a una delle espressioni LF-Tag concesse al principale.

   È possibile specificare le autorizzazioni da concedere ai database corrispondenti, alle tabelle corrispondenti e alle viste corrispondenti.



   In Autorizzazioni database, seleziona le autorizzazioni del database da concedere al principale sui database corrispondenti.

   In Autorizzazioni per le tabelle, seleziona le autorizzazioni per la tabella o la visualizzazione da concedere al principale per le tabelle e le viste corrispondenti.

Puoi anche scegliere tra `Select` le `Describe` `Drop` autorizzazioni della tabella e applicare alle viste le autorizzazioni.

5. Scegli Concessione.

AWS CLI

Puoi utilizzare il metodo AWS Command Line Interface (AWS CLI) e il metodo di controllo degli accessi basato su tag Lake Formation (LF-TBAC) per concedere le autorizzazioni di Lake Formation su database, tabelle e colonne di Data Catalog.

Concessione delle autorizzazioni per il data lake utilizzando il metodo e il metodo LF-TBAC AWS CLI

- Utilizza il comando `grant-permissions`.

  Example

  L'esempio seguente concede all'utente l'espressione LF-Tag "`module=*`" (tutti i valori della chiave LF-Tag). `module` `datalake_user1` Quell'utente avrà l'`CREATE_TABLE`autorizzazione su tutti i database corrispondenti, ovvero i database a cui è stato assegnato il tag LF con la chiave, con qualsiasi valore. `module`

  ```
  aws lakeformation grant-permissions --principal
   DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
  datalake_user1 --permissions "CREATE_TABLE" --resource '{ "LFTagPolicy":
   {"CatalogId":"111122223333","ResourceType":"DATABASE","Expression":
  [{"TagKey":"module","TagValues":["*"]}]}}'
  ```

  Example

  L'esempio successivo concede l'espressione LF-Tag "" all'utente. `(level=director) AND (region=west OR region=south)` `datalake_user1` Quell'utente disporrà delle `DROP` autorizzazioni `SELECT``ALTER`, e con l'opzione grant sulle tabelle corrispondenti, tabelle a cui sono stati assegnati sia (che). `level=director` `region=west` `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

Il prossimo esempio concede l'espressione LF-tag "" all'account 1234-5678-9012. `module=orders` AWS L'amministratore del data lake di quell'account può quindi concedere l'espressione "" ai responsabili del proprio account. `module=orders` Tali responsabili avranno quindi l'`CREATE_TABLE`autorizzazione a far corrispondere i database di proprietà dell'account 1111-2222-3333 e condivisi con l'account 1234-5678-9012 utilizzando il metodo della risorsa denominata o il metodo 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"]}]}}'
```

Scenario di esempio di autorizzazioni

Lo scenario seguente aiuta a dimostrare come è possibile impostare le autorizzazioni per proteggere l'accesso ai dati in. AWS Lake Formation

Shirley è un'amministratore di dati. Vuole creare un data lake per la sua azienda,. AnyCompany Attualmente, tutti i dati sono archiviati in Amazon S3. John è un responsabile marketing e deve accedere per iscritto alle informazioni sugli acquisti dei clienti (contenute in`s3://customerPurchases`). Un analista di marketing, Diego, si unisce a John quest'estate. John deve poter concedere a Diego l'accesso per eseguire interrogazioni sui dati senza coinvolgere Shirley.

Mateo, del reparto finanziario, ha bisogno di accedere ai dati contabili interrogativi (ad esempio,). `s3://transactions` Vuole interrogare i dati delle transazioni nelle tabelle di un database (`Finance_DB`) utilizzato dal team finanziario. Il suo manager, Arnav, può dargli accesso a. `Finance_DB` Sebbene non dovrebbe essere in grado di modificare i dati contabili, deve poter convertire i dati in un formato (schema) adatto alla previsione. Questi dati verranno archiviati in un bucket separato (`s3://financeForecasts`) che potrà modificare.

Per riassumere:

- Shirley è l'amministratore del data lake.
- John richiede `CREATE_DATABASE` `CREATE_TABLE` l'autorizzazione per creare nuovi database e tabelle nel Data Catalog.
- John richiede `SELECT` anche le autorizzazioni e `DELETE` le autorizzazioni per le tabelle che crea. `INSERT`
- Diego richiede `SELECT` l'autorizzazione sul tavolo per eseguire le interrogazioni.

I dipendenti di AnyCompany eseguono le seguenti azioni per impostare le autorizzazioni. Le operazioni API mostrate in questo scenario mostrano una sintassi semplificata per motivi di chiarezza.

1. Shirley registra il percorso Amazon S3 contenente le informazioni sugli acquisti dei clienti con Lake Formation.

   ```
   RegisterResource(ResourcePath("s3://customerPurchases"), false, Role_ARN )
   ```

2. Shirley concede a John l'accesso al percorso Amazon S3 contenente le informazioni di acquisto dei clienti.

   ```
   GrantPermissions(John, S3Location("s3://customerPurchases"),
    [DATA_LOCATION_ACCESS]) )
   ```

3. Shirley concede a John il permesso di creare database.

   ```
   GrantPermissions(John, catalog, [CREATE_DATABASE])
   ```

4. John crea il database. `John_DB` John ha automaticamente `CREATE_TABLE` l'autorizzazione per quel database perché lo ha creato.

   ```
   CreateDatabase(John_DB)
   ```

5. John crea la tabella `John_Table` che punta a. `s3://customerPurchases` Poiché ha creato la tabella, dispone di tutte le autorizzazioni e può concedere le autorizzazioni necessarie.

   ```
   CreateTable(John_DB, John_Table)
   ```

6. John consente al suo analista, Diego, di accedere alla tabella. `John_Table`

```
GrantPermissions(Diego, John_Table, [SELECT])
```

7. John consente al suo analista, Diego, di accedere a. `s3://customerPurchases/London/` Poiché Shirley è già registrata`s3://customerPurchases`, le sue sottocartelle sono registrate con Lake Formation.

```
GrantDataLakePrivileges( 123456789012/datalake, Diego, [DATA_LOCATION_ACCESS], [],
S3Location("s3://customerPurchases/London/") )
```

8. John consente al suo analista, Diego, di creare tabelle nel database. `John_DB`

```
GrantDataLakePrivileges( 123456789012/datalake, Diego, John_DB, [CREATE_TABLE],
[] )
```

9. Diego crea una tabella `John_DB` in `s3://customerPurchases/London/` e ottiene automaticamente`ALTER`,, `DROP` `SELECT``INSERT`, e le `DELETE` autorizzazioni.

```
CreateTable( 123456789012/datalake, John_DB, Diego_Table )
```

Filtraggio dei dati e sicurezza a livello di cella in Lake Formation

Quando concedi le autorizzazioni di Lake Formation su una tabella Data Catalog, puoi includere specifiche di filtraggio dei dati per limitare l'accesso a determinati dati nei risultati delle query e nei motori integrati con Lake Formation. Lake Formation utilizza il filtraggio dei dati per ottenere la sicurezza a livello di colonna, la sicurezza a livello di riga e la sicurezza a livello di cella. Puoi definire e applicare filtri di dati sulle colonne nidificate se i dati di origine contengono strutture nidificate.

Argomenti

Panoramica del filtraggio dei dati

Con le funzionalità di filtraggio dei dati di Lake Formation, puoi implementare i seguenti livelli di sicurezza dei dati.

Sicurezza a livello di colonna

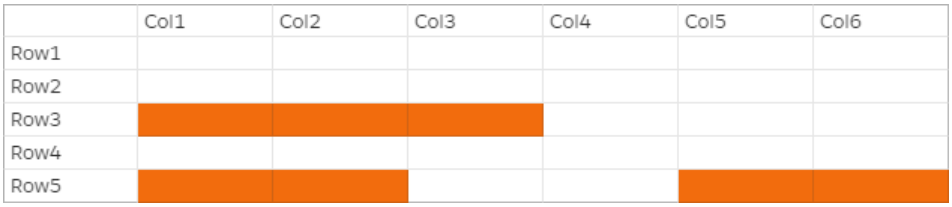
La concessione delle autorizzazioni su una tabella del Catalogo dati con sicurezza a livello di colonna (filtro delle colonne) consente agli utenti di visualizzare solo colonne specifiche e colonne nidificate a cui hanno accesso nella tabella. Prendiamo in considerazione una `persons` tabella utilizzata in più applicazioni per una grande società di comunicazioni multiregionale. La concessione di autorizzazioni per le tabelle di Data Catalog con filtro a colonne può impedire agli utenti che non lavorano nel reparto risorse umane di visualizzare informazioni di identificazione personale (PII) come il numero di previdenza sociale o la data di nascita. È inoltre possibile definire politiche di sicurezza e concedere l'accesso solo a sottostrutture parziali di colonne annidate.

Sicurezza a livello di riga

La concessione delle autorizzazioni su una tabella del Catalogo dati con sicurezza a livello di riga (filtro di riga) consente agli utenti di visualizzare solo righe di dati specifiche a cui hanno accesso nella tabella. Il filtro si basa sui valori di una o più colonne. È possibile includere strutture di colonne annidate quando si definiscono espressioni di filtro di riga. Ad esempio, se diversi uffici regionali della società di comunicazioni hanno i propri reparti delle risorse umane, è possibile limitare i record relativi alle persone che i dipendenti delle risorse umane possono vedere ai soli record relativi ai dipendenti della propria regione.

Sicurezza a livello di cella

La sicurezza a livello di cella combina il filtraggio di righe e il filtraggio di colonne per un modello di autorizzazioni altamente flessibile. Se si visualizzano le righe e le colonne di una tabella come griglia, utilizzando la sicurezza a livello di cella, è possibile limitare l'accesso ai singoli elementi (celle) della griglia in qualsiasi punto delle due dimensioni. In altre parole, è possibile limitare l'accesso a diverse colonne a seconda della riga. Ciò è illustrato dal diagramma seguente, in cui le colonne con restrizioni sono ombreggiate.



Continuando l'esempio della tabella delle persone, puoi creare un filtro dati a livello di cella che limita l'accesso alla colonna dell'indirizzo se la riga ha la colonna del paese impostata su «UK», ma consente l'accesso alla colonna dell'indirizzo se la riga ha la colonna del paese impostata su «US».

I filtri si applicano solo alle operazioni di lettura. Pertanto, puoi concedere solo l'autorizzazione `SELECT` Lake Formation con filtri.

Sicurezza a livello di cella sulle colonne annidate

Lake Formation consente di definire e applicare filtri di dati con sicurezza a livello di cella su colonne annidate. Tuttavia, i motori analitici integrati come Amazon Athena, Amazon EMR e Amazon Redshift Spectrum supportano l'esecuzione di query su tabelle nidificate gestite da Lake Formation con sicurezza a livello di riga e colonna.

Per le limitazioni, consulta Limitazioni del filtraggio dei dati.

Filtri di dati in Lake Formation

Puoi implementare la sicurezza a livello di colonna, a livello di riga e a livello di cella creando filtri di dati. Seleziona un filtro dati quando concedi l'autorizzazione `SELECT` Lake Formation sulle tabelle. Se la tabella contiene strutture di colonne nidificate, è possibile definire un filtro dati includendo o escludendo le colonne secondarie e definire espressioni di filtro a livello di riga sugli attributi nidificati.

Ogni filtro di dati appartiene a una tabella specifica del Data Catalog. Un filtro dati include le seguenti informazioni:

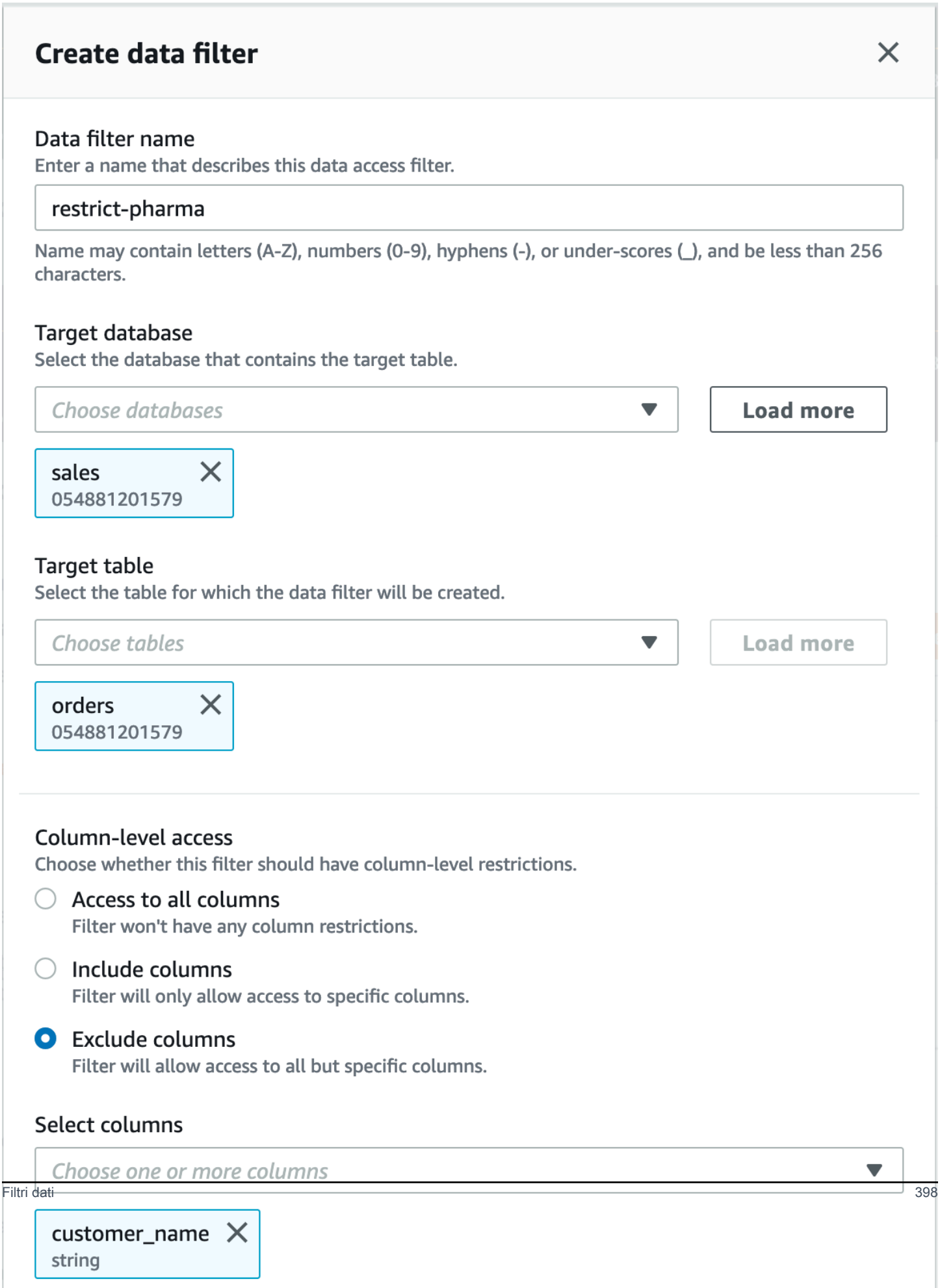
- Nome del filtro
- Gli ID del catalogo della tabella associata al filtro
- Nome tabella
- Nome del database che contiene la tabella
- Specificazione delle colonne: un elenco di colonne e colonne annidate (con `struct` tipi di dati) da includere o escludere nei risultati della query.
- Espressione di filtro di riga: un'espressione che specifica le righe da includere nei risultati della query. Con alcune restrizioni, l'espressione ha la sintassi di una `WHERE` clausola nel linguaggio PartiQL. Per specificare tutte le righe, scegli Accesso a tutte le righe in Accesso a livello di riga nella console o Utilizza nelle chiamate API. `AllRowsWildcard`

Per ulteriori informazioni su ciò che è supportato nelle espressioni di filtro di riga, consulta. Supporto PartiQL nelle espressioni di filtro di riga

Il livello di filtraggio ottenuto dipende da come si popola il filtro dati.

- Quando specifichi il carattere jolly "tutte le colonne" e fornisci un'espressione di filtro di riga, stabilisci solo la sicurezza a livello di riga (filtraggio di riga).
- Quando si includono o si escludono colonne specifiche e colonne nidificate e si specifica «tutte le righe» utilizzando il carattere jolly per tutte le righe, si stabilisce solo la sicurezza a livello di colonna (filtraggio delle colonne).
- Quando includi o escludi colonne specifiche e fornisci anche un'espressione di filtro di riga, stabilisci la sicurezza a livello di cella (filtraggio delle celle).

La seguente schermata della console Lake Formation mostra un filtro dati che esegue il filtraggio a livello di cella. Per le interrogazioni sulla `orders` tabella, limita l'accesso alla `customer_name` colonna e i risultati della query restituiscono solo le righe in cui la colonna contiene «pharma». `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

Nota l'uso di virgolette singole per racchiudere la stringa letterale,. `'pharma'`

Puoi utilizzare la console Lake Formation per creare questo filtro di dati oppure puoi fornire il seguente oggetto di richiesta all'operazione `CreateDataCellsFilter` API.

```
{
     "Name": "restrict-pharma",
     "DatabaseName": "sales",
     "TableName": "orders",
     "TableCatalogId": "111122223333",
     "RowFilter": {"FilterExpression": "product_type='pharma'"},
     "ColumnWildcard": {
         "ExcludedColumnNames": ["customer_name"]
     }
}
```

Puoi creare tutti i filtri di dati di cui hai bisogno per una tabella. A tal fine, è necessaria l'`SELECT`autorizzazione con l'opzione di concessione su una tabella. Per impostazione predefinita, gli amministratori di Data Lake dispongono dell'autorizzazione per creare filtri di dati su tutte le tabelle di quell'account. In genere si utilizza solo un sottoinsieme dei possibili filtri di dati quando si concedono le autorizzazioni sulla tabella a un responsabile. Ad esempio, è possibile creare un secondo filtro di dati per la `orders` tabella che è un row-security-only filtro di dati. Facendo riferimento alla schermata precedente, è possibile scegliere l'opzione Accesso a tutte le colonne e includere un'espressione di filtro di riga di. `product_type<>pharma` Il nome di questo filtro di dati potrebbe essere. `no-pharma` Limita l'accesso a tutte le righe la cui `product_type` colonna è impostata su «pharma».

L'oggetto di richiesta per l'operazione `CreateDataCellsFilter` API per questo filtro di dati è il seguente.

```
{
     "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"]
}
```

È quindi possibile concedere la licenza `SELECT` sulla `orders` tabella con il filtro `restrict-pharma` dati a un utente amministrativo e `SELECT` sulla `orders` tabella con il filtro `no-pharma` dati a utenti non amministrativi. Agli utenti del settore sanitario, potresti `SELECT` concedere alla `orders` tabella l'accesso completo a tutte le righe e le colonne (nessun filtro per i dati) o magari aggiungere un altro filtro di dati che limiti l'accesso alle informazioni sui prezzi.

È possibile includere o escludere colonne annidate quando si specifica la sicurezza a livello di colonna e di riga all'interno di un filtro dati. Nell'esempio seguente, l'accesso al `product.offer` campo viene specificato utilizzando nomi di colonna qualificati (racchiusi tra virgolette doppie). Questo è importante per i campi annidati per evitare che si verifichino errori quando i nomi delle colonne contengono caratteri speciali e per mantenere la compatibilità con le versioni precedenti delle definizioni di sicurezza a livello di colonna.

```
{
     "Name": "example_dcf",
     "DatabaseName": "example_db",
     "TableName": "example_table",
     "TableCatalogId": "111122223333",
     "RowFilter": { "FilterExpression": "customer.customerName <> 'John'" },
     "ColumnNames": ["customer", "\"product\".\"offer\""]
}
```

Consulta anche

- Gestione dei filtri di dati

Supporto PartiQL nelle espressioni di filtro di riga

È possibile creare espressioni di filtro di riga utilizzando un sottoinsieme di tipi di dati, operatori e aggregazioni PartiQL. Lake Formation non consente alcuna funzione PartiQL definita dall'utente o standard nell'espressione del filtro. È possibile utilizzare gli operatori di confronto per confrontare le colonne con costanti (ad esempio,`views >= 10000`), ma non è possibile confrontare le colonne con altre colonne.

Un'espressione di filtro Row può essere un'espressione semplice o un'espressione composita. La lunghezza totale dell'espressione deve essere inferiore a 2048 caratteri.

Espressione semplice

Un'espressione semplice avrà il seguente formato: `<column name > <comparison operator ><value >`

- Nome della colonna

  Può essere una colonna di dati di primo livello, una colonna di partizione o una colonna nidificata presente nello schema della tabella e deve appartenere alle colonne Tipi di dati supportati elencate di seguito.

- Operatore di confronto

  Gli operatori supportati sono i seguenti: `=, >, <, >=, <=, <>,!=, BETWEEN, IN, LIKE, NOT, IS [NOT] NULL`

- Tutti i confronti tra stringhe e le corrispondenze di `LIKE` modelli fanno distinzione tra maiuscole e minuscole. Non è possibile utilizzare l'operatore IS [NOT] NULL sulle colonne di partizione.
- Valore della colonna

  Il valore della colonna deve corrispondere al tipo di dati del nome della colonna.

Espressione composita

Un'espressione composita avrà il formato:`( <simple expression >) <AND/OR >(<simple expression >)`. Le espressioni composite possono essere ulteriormente combinate utilizzando operatori logici`AND/OR`.

Tipi di dati supportati

I filtri di riga che fanno riferimento a una AWS Glue Data Catalog tabella che contiene un tipo di dati non supportato genereranno un errore. Di seguito sono riportati i tipi di dati supportati per le colonne e le costanti della tabella, che sono mappati ai Amazon Redshift tipi di dati:

- `STRING, CHAR, VARCHAR`
- `INT, LONG, BIGINT, FLOAT, DECIMAL, DOUBLE`
- `BOOLEAN`
- `STRUCT`

Per ulteriori informazioni sui tipi di dati in Amazon Redshift, consulta Tipi di dati nella Amazon Redshift Database Developer Guide.

Espressioni di filtro di riga

Example

Di seguito sono riportati alcuni esempi di espressioni di filtro di riga valide per una tabella con colonne: `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

Di seguito sono riportati alcuni esempi validi di espressioni di filtro di riga per una tabella con colonne annidate: `year > 2010 and customer.customerId <> 1`

Non è necessario fare riferimento ai campi nidificati nelle colonne di partizione quando si definiscono espressioni nidificate a livello di riga.

Le costanti di stringa devono essere racchiuse tra virgolette singole.

Parole chiave riservate

Se l'espressione del filtro di riga contiene parole chiave PartiQL, riceverai un errore di analisi poiché i nomi delle colonne potrebbero essere in conflitto con le parole chiave. Quando ciò accade, evita i nomi delle colonne usando le virgolette doppie. Alcuni esempi di parole chiave riservate sono «first», «last», «asc», «missing». Vedi la specifica PartiQL per un elenco di parole chiave riservate.

Riferimento PartiQL

Per ulteriori informazioni su PartiQL, vedere. https://partiql.org/

Note e restrizioni per il filtraggio a livello di colonna

Esistono tre modi per specificare il filtraggio delle colonne:

- Utilizzando i filtri di dati, come descritto in precedenza.
- Utilizzando un semplice filtraggio a colonne o un filtraggio a colonne annidate.
- Utilizzando i TAG.

Il semplice filtraggio delle colonne specifica solo un elenco di colonne da includere o escludere. Sia la console di Lake Formation che l'API AWS CLI supportano il semplice filtraggio delle colonne. Per vedere un esempio, consulta Grant with Simple Column Filtering.

Le seguenti note e restrizioni si applicano al filtraggio delle colonne:

- AWS Gluel job ETL supportano il filtraggio delle colonne solo utilizzando filtri di dati (sicurezza a livello di cella). Il processo ha esito negativo se viene applicato un semplice filtro a colonne a qualsiasi tabella a cui il lavoro fa riferimento. Se desideri filtrare solo le colonne, concedi l'accesso alle tabelle utilizzando i filtri di dati e inserisci `true` l'espressione del filtro di riga nella console oppure utilizzala `AllRowsWildcard` nelle tue chiamate API.
- Per concedere `SELECT` con l'opzione grant e il filtraggio delle colonne, devi utilizzare un elenco di inclusione, non un elenco di esclusione. Senza l'opzione di concessione, è possibile utilizzare elenchi di inclusione o esclusione.
- Per concedere `SELECT` su una tabella con filtro a colonne, devi aver ottenuto l'autorizzazione `SELECT` sulla tabella con l'opzione di concessione e senza alcuna restrizione di riga. Devi avere accesso a tutte le righe.
- Se `SELECT` concedi l'opzione di concessione e il filtraggio delle colonne a un principale del tuo account, tale principale deve specificare il filtraggio delle colonne per le stesse colonne o per un sottoinsieme delle colonne concesse quando concede a un altro principale. Se concedi l'opzione `SELECT` di concessione e il filtraggio delle colonne a un account esterno, l'amministratore del data lake dell'account esterno può concedere `SELECT` tutte le colonne a un altro responsabile del proprio account. Tuttavia, anche se `SELECT` su tutte le colonne, tale principale avrà visibilità solo sulle colonne concesse all'account esterno.
- Non è possibile applicare il filtraggio delle colonne sulle chiavi di partizione.
- A un principale con l'`SELECT`autorizzazione su un sottoinsieme di colonne di una tabella non può essere concessa l'`INSERT`autorizzazione`ALTER`,, `DROP``DELETE`, o per quella tabella. Per un'entità principale con l'`INSERT`autorizzazione`ALTER`, `DROP``DELETE`, o su una tabella, se si concede l'`SELECT`autorizzazione con il filtro delle colonne, l'autorizzazione non ha alcun effetto.

Le seguenti note e restrizioni si applicano al filtraggio delle colonne annidate:

- È possibile includere o escludere cinque livelli di campi nidificati in un filtro dati.

  Example

  Col1.Col1_1.Col1_1_1.Col1_1_1_1.Col1_1_1_1_1_1
- Non è possibile applicare il filtraggio delle colonne ai campi annidati all'interno delle colonne delle partizioni.
- Se lo schema della tabella contiene un nome di colonna di primo livello («cliente»).» address») che presenta lo stesso schema di rappresentazione di un campo nidificato all'interno di un filtro dati (una colonna nidificata con un nome di colonna di primo livello `customer` e un nome di campo nidificato `address` viene specificata come `"customer"."address"` in un filtro dati), non è possibile specificare in modo esplicito l'accesso alla colonna di primo livello o al campo nidificato perché entrambi sono rappresentati utilizzando lo stesso modello negli elenchi di inclusione/esclusione. Questo è ambiguo e Lake Formation non può risolvere se stai specificando la colonna di primo livello o il campo annidato.
- Se una colonna di primo livello o un campo nidificato contiene una doppia virgoletta all'interno del nome, devi includere una seconda virgoletta quando specifichi l'accesso a un campo nidificato all'interno dell'elenco di inclusione ed esclusione di un filtro di celle di dati.

  Example

  Esempio di nome di colonna nidificata con virgolette doppie: `a.b.double"quote`

  Example

  Esempio di rappresentazione di una colonna annidata all'interno di un filtro dati: `"a"."b"."double""quote"`

Autorizzazioni necessarie per eseguire query su tabelle con filtraggio a livello di cella

Le seguenti autorizzazioni AWS Identity and Access Management (IAM) sono necessarie per eseguire query su tabelle con filtraggio a livello di cella.

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

```
            "Effect": "Allow",
            "Action": [
                "lakeformation:StartQueryPlanning",
                "lakeformation:GetQueryState",
                "lakeformation:GetWorkUnits",
                "lakeformation:GetWorkUnitResults"
            ],
            "Resource": "*"
        }
    ]
}
```

Per ulteriori informazioni sui permessi di Lake Formation, vedereRiferimento ai personaggi di Lake Formation e alle autorizzazioni IAM.

Gestione dei filtri di dati

Per implementare la sicurezza a livello di colonna, riga e cella, puoi creare e gestire filtri di dati. Ogni filtro di dati appartiene a una tabella Data Catalog. È possibile creare più filtri di dati per una tabella e quindi utilizzarne uno o più per concedere le autorizzazioni sulla tabella. È inoltre possibile definire e applicare filtri di dati su colonne nidificate con `struct` tipi di dati che consentono agli utenti di accedere solo alle sottostrutture delle colonne nidificate.

È necessaria `SELECT` l'autorizzazione con l'opzione di concessione per creare o visualizzare un filtro dati. Per consentire ai responsabili del tuo account di visualizzare e utilizzare un filtro dati, puoi concedere l'`DESCRIBE`autorizzazione su di esso.

> **Note**
>
> Lake Formation non supporta la concessione `Describe` dell'autorizzazione su un filtro dati, che viene condiviso da un altro account.

Puoi gestire i filtri dei dati utilizzando la AWS Lake Formation console, l'API o AWS Command Line Interface (AWS CLI).

Per informazioni sui filtri di dati, consulta Filtri di dati in Lake Formation

Creazione di un filtro dati

È possibile creare uno o più filtri di dati per ogni tabella del Data Catalog.

Per creare un filtro dati per una tabella Data Catalog (console)

1. Aprire la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.

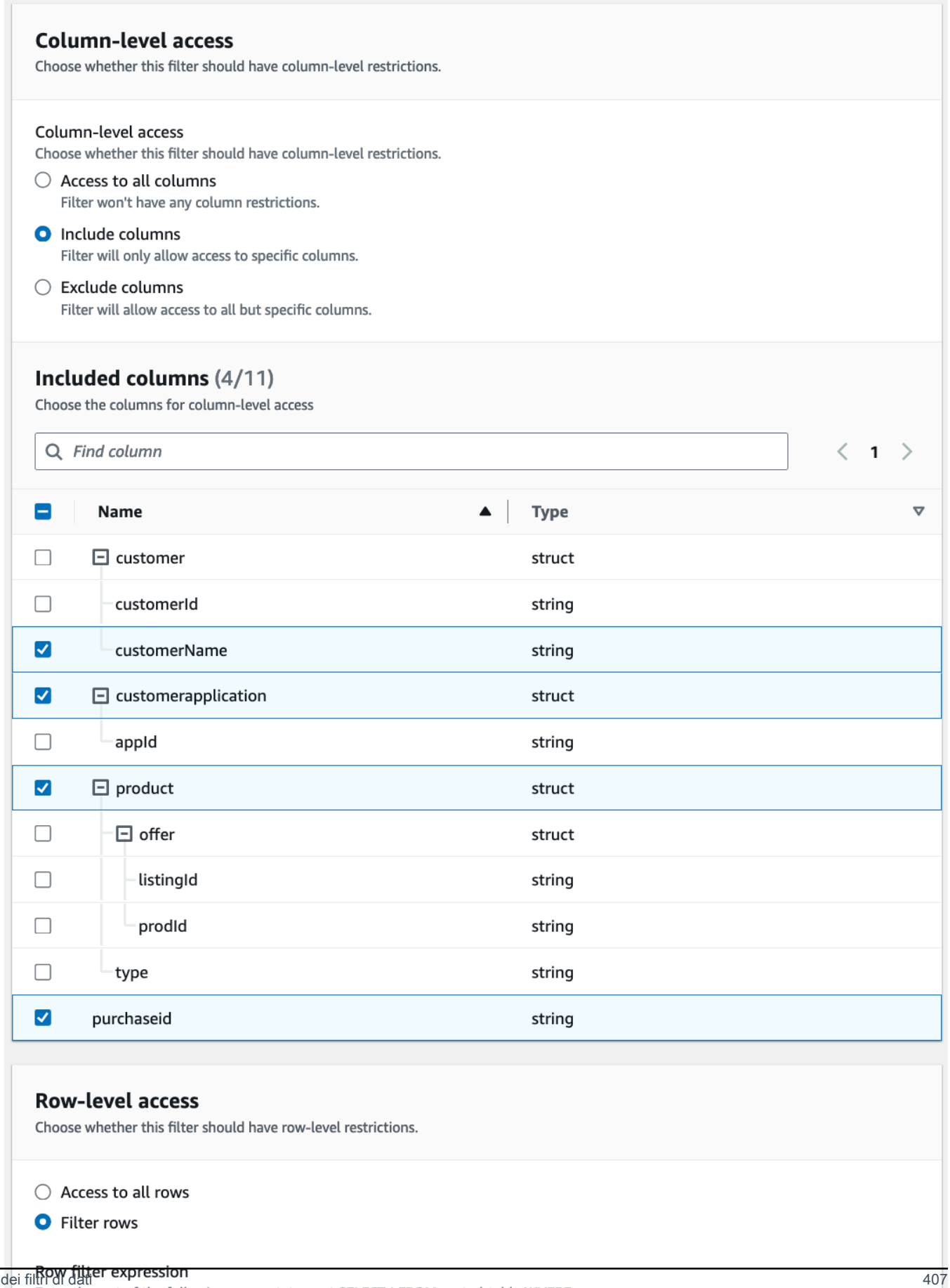
   Accedi come amministratore del data lake, proprietario della tabella di destinazione o principale che dispone dell'autorizzazione Lake Formation sulla tabella di destinazione.
2. Nel riquadro di navigazione, in Catalogo dati, scegli Filtri dati.
3. Nella pagina Filtri dati, scegli Crea nuovo filtro.
4. Nella finestra di dialogo Crea filtro dati, inserisci le seguenti informazioni:

   - Nome del filtro dati
   - Database di destinazione: specifica il database che contiene la tabella.
   - Tabella di destinazione
   - Accesso a livello di colonna: lascia questa impostazione su Accesso a tutte le colonne per specificare solo il filtraggio delle righe. Scegliete Includi colonne o Escludi colonne per specificare il filtraggio di colonne o celle, quindi specificate le colonne da includere o escludere.

     Colonne nidificate: se applichi il filtro a una tabella che contiene colonne nidificate, puoi specificare in modo esplicito le sottostrutture delle colonne struct nidificate all'interno di un filtro dati.

     Quando concedi l'autorizzazione SELECT a un principale su questo filer, il principale che esegue la seguente query vedrà solo i dati relativi e non. `customer.customerName` `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'

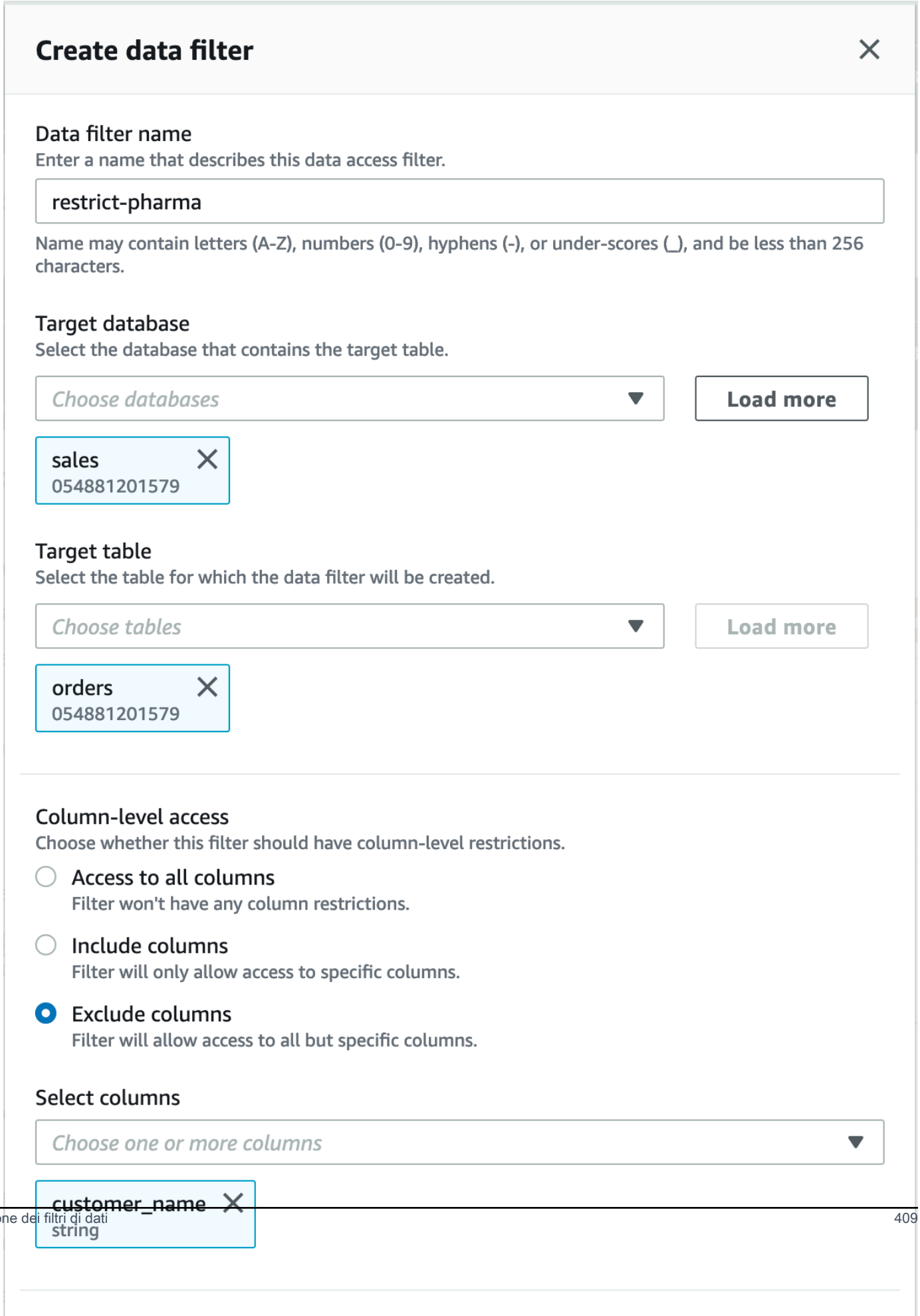
Quando concedi le autorizzazioni alla `customer` colonna, il principale riceve l'accesso alla colonna e ai campi annidati sotto la colonna (`customerNamee`). `customerID`

- Espressione di filtro di riga: immettere un'espressione di filtro per specificare il filtraggio di righe o celle. Per i tipi di dati e gli operatori supportati, vedereSupporto PartiQL nelle espressioni di filtro di riga. Scegli Accesso a tutte le righe per concedere l'accesso a tutti.

  È possibile includere strutture di colonne parziali provenienti da colonne nidificate in un'espressione di filtro di riga per filtrare le righe che contengono un valore specifico.

  Quando a un'entità vengono concesse le autorizzazioni per una tabella con un'espressione `Select * from example_nestedtable where customer.customerName <>'John'` di filtro di riga e l'accesso a livello di colonna è impostato su Accesso a tutte le colonne, i risultati della query mostrano solo le righe il cui risultato è true. `customerName <>'John'`

La schermata seguente mostra un filtro dati che implementa il filtraggio delle celle. Nelle interrogazioni sulla `orders` tabella, nega l'accesso alla `customer_name` colonna e mostra solo le righe che contengono «pharma» nella colonna. `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. Scegli Create Filter (Crea filtro).

Per creare un filtro dati con politiche di filtro cellulare su un campo annidato

Questa sezione utilizza lo schema di esempio seguente per mostrare come creare un filtro di celle di dati:

```
[
    { 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. Nella pagina Crea un filtro dati, inserisci un nome per il filtro dati.
2. Successivamente, utilizza il menu a discesa per scegliere il nome del database e il nome della tabella.
3. Nella sezione Accesso a livello di colonna, scegli Colonne incluse e seleziona una colonna nidificata (). `customer.customerName`
4. Nella sezione Accesso a livello di riga, scegli l'opzione Accesso a tutte le righe.
5. Scegli Create Filter (Crea filtro).

   Quando concedi `SELECT` l'autorizzazione per questo filtro, il principale ottiene l'accesso a tutte le righe della `customerName` colonna.
6. Successivamente, definisci un altro filtro dati per lo stesso database/tabella.
7. Nella sezione Accesso a livello di colonna, scegli Colonne incluse e seleziona un'altra colonna nidificata (). `customer.customerid`
8. Nella sezione Accesso a livello di riga, scegliete Filtra righe e immettete un'espressione di filtro di riga (). `customer.customerid <> 5`
9. Scegli Create Filter (Crea filtro).

   Quando concedete `SELECT` l'autorizzazione per questo filtro, il principale riceve l'accesso a tutte le righe e ai `customerId` campi ad eccezione della cella in cui il valore è 5 nella `customerId` colonna. `customerName`

Concessione delle autorizzazioni per il filtro dei dati

Puoi concedere le autorizzazioni`SELECT`, `DESCRIBE` e `DROP` Lake Formation sui filtri di dati ai responsabili.

All'inizio, solo tu puoi visualizzare i filtri di dati che crei per una tabella. Per consentire a un altro principale di visualizzare un filtro di dati e concedere le autorizzazioni di Data Catalog con il filtro dati, devi:

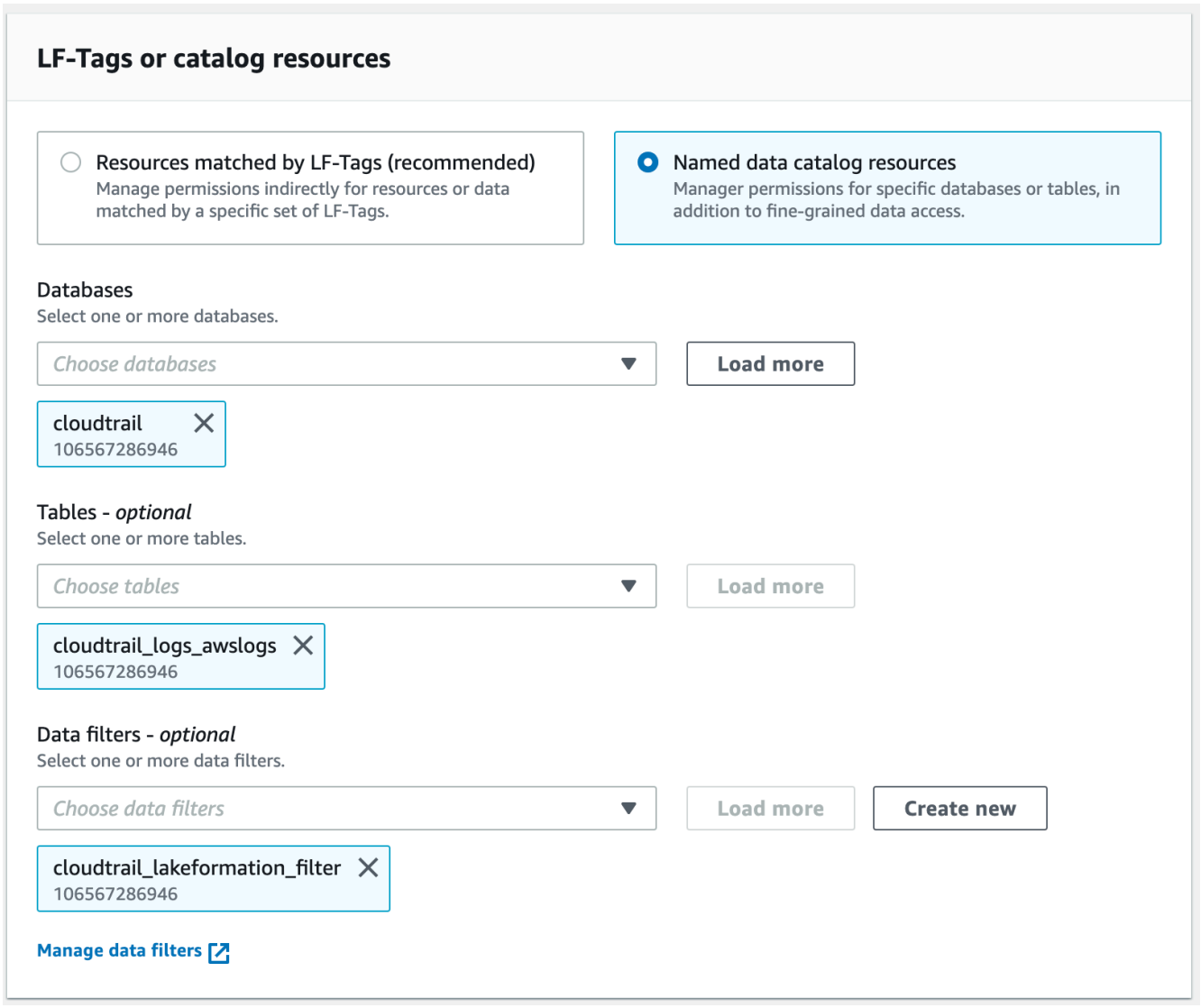
- `SELECT`Concedi una tabella al principale con l'opzione di concessione e applica il filtro dati alla concessione.
- Concedi l'`DROP`autorizzazione `DESCRIBE` o l'autorizzazione sul filtro dati al responsabile.

È possibile concedere l'`SELECT`autorizzazione a un AWS account esterno. Un amministratore del data lake di quell'account può quindi concedere tale autorizzazione ad altri responsabili dell'account. Quando si concede a un account esterno, è necessario includere l'opzione di concessione in modo che l'amministratore dell'account esterno possa trasferire ulteriormente l'autorizzazione ad altri utenti del proprio account. Quando concedi la concessione a un titolare del tuo account, la concessione con l'opzione di concessione è facoltativa.

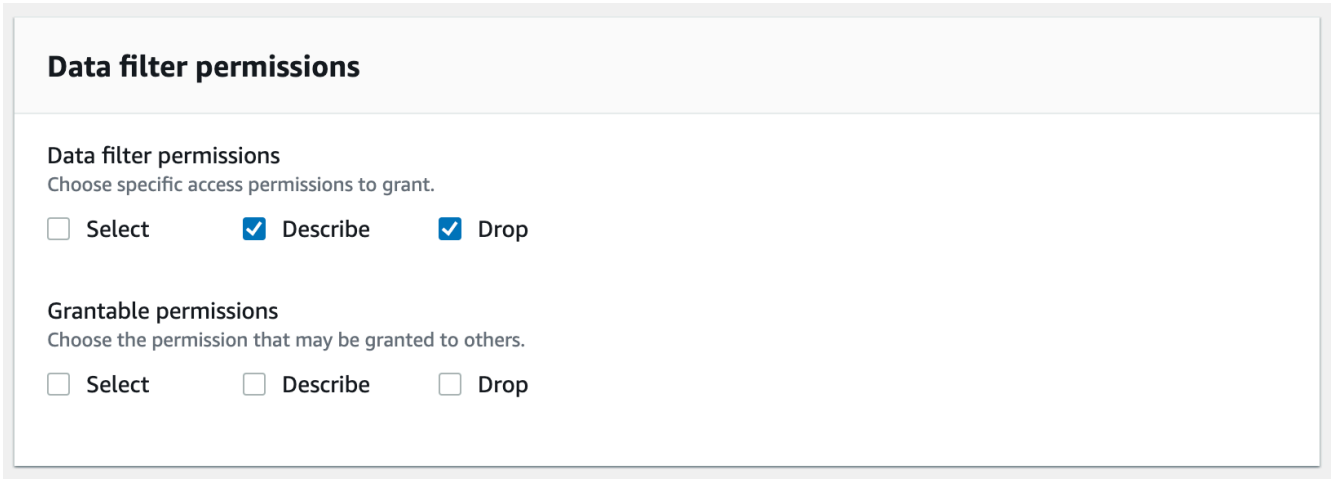
Puoi concedere e revocare le autorizzazioni sui filtri di dati utilizzando la AWS Lake Formation console, l'API o (). AWS Command Line Interface AWS CLI

Console

1. Accedi AWS Management Console e apri la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.
2. Nel pannello di navigazione, sotto Autorizzazioni, scegli Autorizzazioni Data lake.
3. Nella pagina Autorizzazioni, nella sezione Autorizzazioni dati, scegli Concedi.
4. Nella pagina Concedi le autorizzazioni per i dati, scegli i principali a cui concedere le autorizzazioni.
5. Nella sezione LF-Tags o risorse del catalogo, scegliete Risorse del catalogo dati denominate. Scegliete quindi il database, la tabella e il filtro dati per i quali desiderate concedere le autorizzazioni.



6. Nella sezione Autorizzazioni del filtro dati, scegli le autorizzazioni che desideri concedere ai principali selezionati.



AWS CLI

- Inserisci un comando. `grant-permissions` Specificate `DataCellsFilter` per l'`resource`argomento e specificate `DESCRIBE` o `DROP` per l'`Permissions`argomento e, facoltativamente, per l'`PermissionsWithGrantOption`argomento.

  L'esempio seguente concede l'opzione `DESCRIBE` di concessione all'utente `datalake_user1` sul filtro dati`restrict-pharma`, che appartiene alla `orders` tabella del `sales` database dell'AWSaccount 1111-2222-3333.

  ```
  aws lakeformation grant-permissions --cli-input-json file://grant-params.json
  ```

  Di seguito è riportato il contenuto del file. `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"]
  }
  ```

Concessione delle autorizzazioni relative ai dati fornite dai filtri di dati

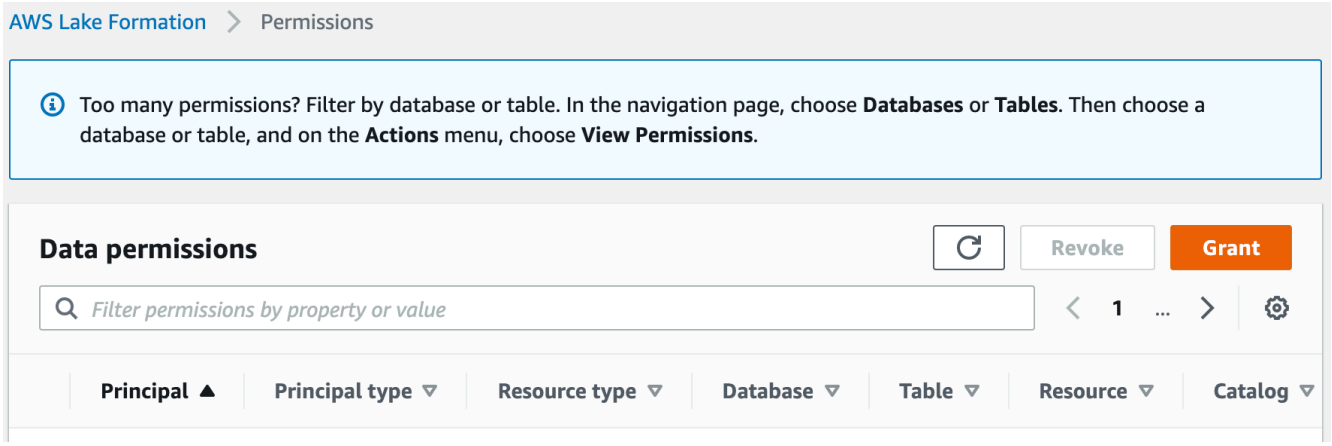
I filtri di dati rappresentano un sottoinsieme di dati all'interno di una tabella. Per fornire l'accesso ai dati ai principali, è necessario concedere `SELECT` le autorizzazioni a tali responsabili. Con questa autorizzazione i responsabili possono:

- Visualizza il nome effettivo della tabella nell'elenco delle tabelle condivise con il loro account.
- Crea filtri di dati sulla tabella condivisa e concedi le autorizzazioni ai relativi utenti su tali filtri di dati.

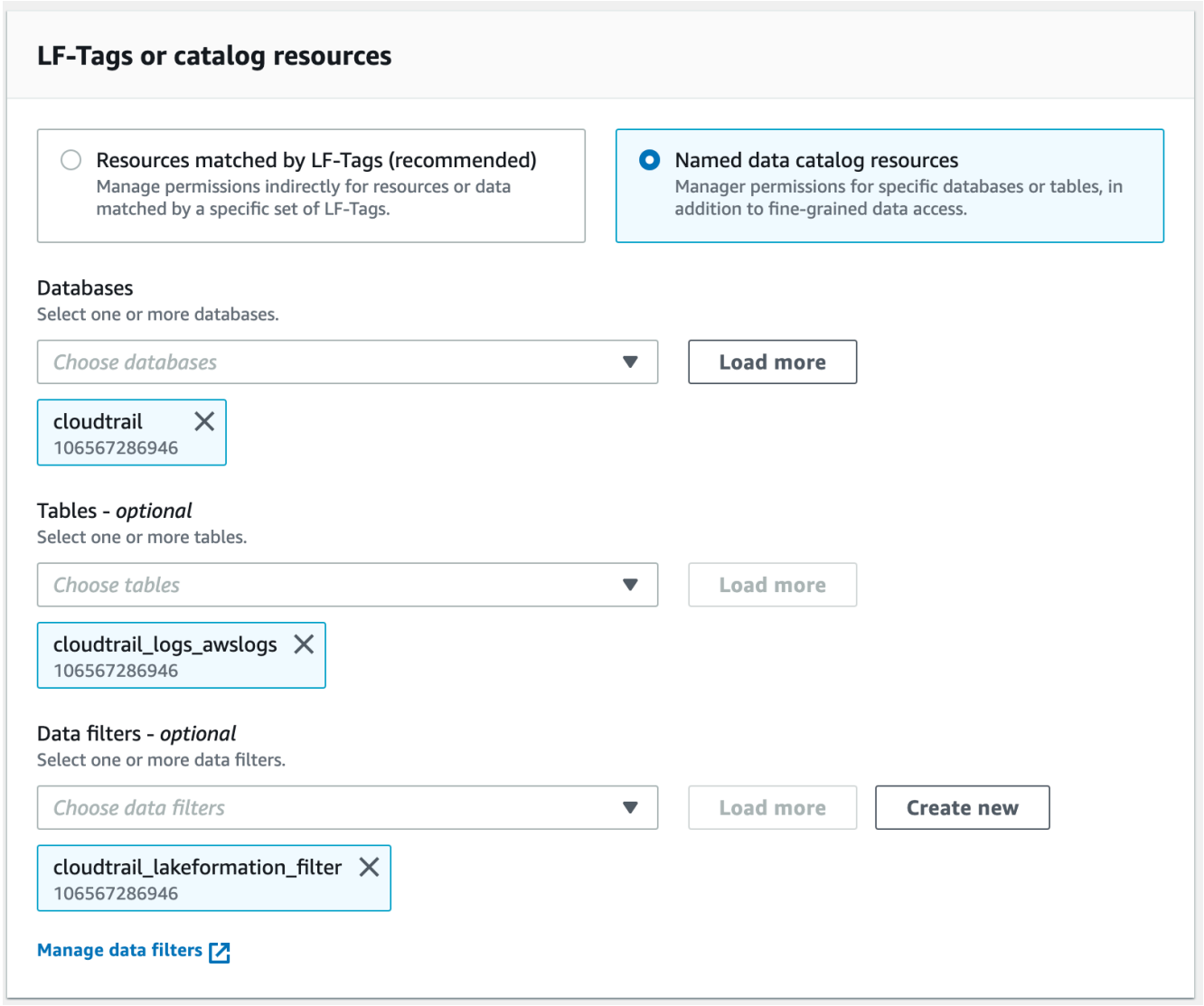
Console

Per concedere le autorizzazioni SELECT

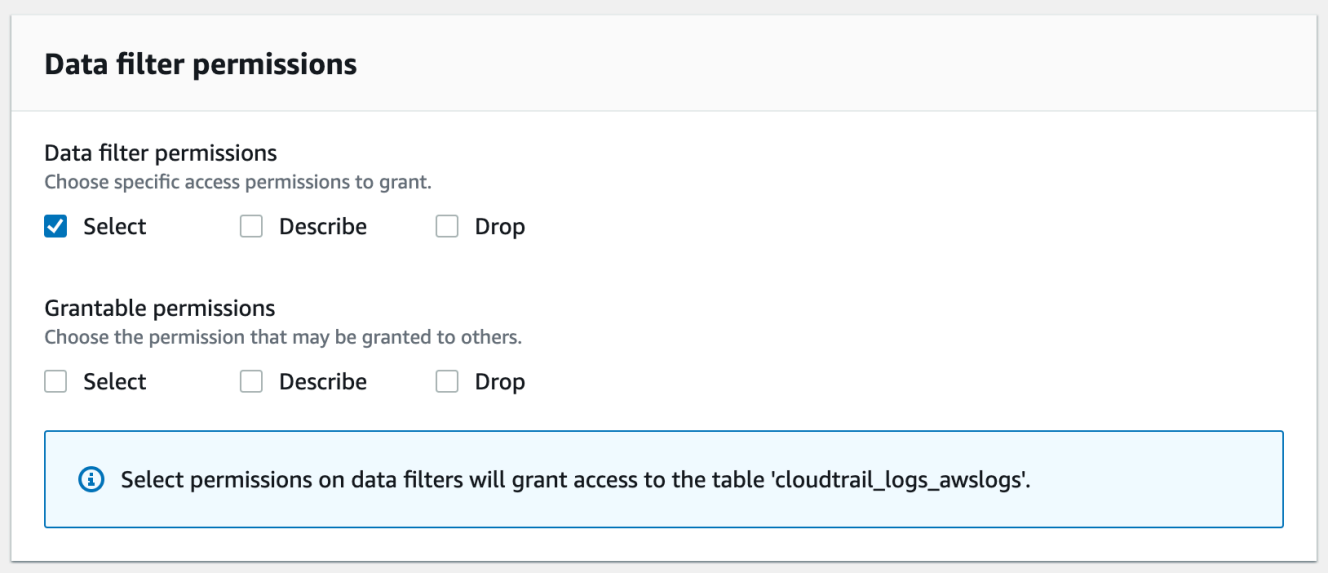
1. Vai alla pagina Autorizzazioni nella console di Lake Formation, quindi scegli Concedi.



2. Seleziona i principali a cui desideri fornire l'accesso e seleziona Named data catalog resources.



3. Per fornire l'accesso ai dati rappresentati dal filtro, scegli Seleziona in Autorizzazioni del filtro dati.



CLI

Inserisci `grant-permissions` un comando. `DataCellsFilter`Specificate l'argomento risorsa e specificate `SELECT` l'argomento Autorizzazioni.

L'esempio seguente concede `SELECT` l'opzione grant all'utente `datalake_user1` sul filtro dati`restrict-pharma`, che appartiene alla `orders` tabella del `sales` database in. Account AWS `1111-2222-3333`

```
aws lakeformation grant-permissions --cli-input-json file://grant-params.json
```

Di seguito sono riportati i contenuti del file`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"]
}
```

Visualizzazione dei filtri di dati

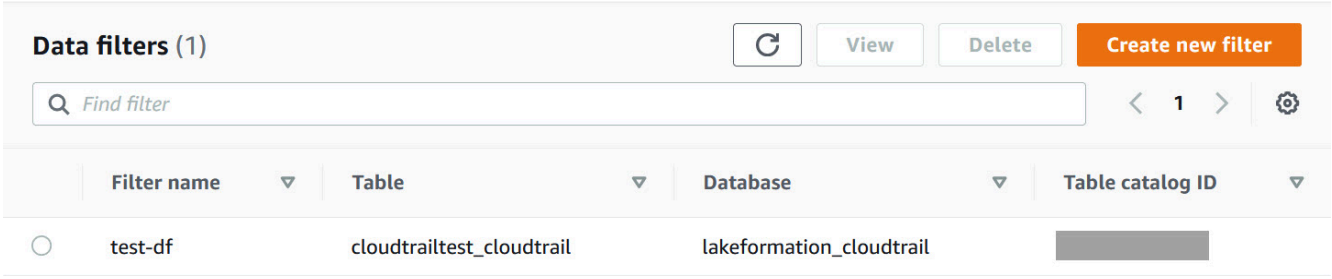
Puoi utilizzare la console Lake Formation o AWS CLI l'API Lake Formation per visualizzare i filtri dei dati.

Per visualizzare i filtri di dati, devi essere un amministratore di Data Lake o disporre delle autorizzazioni richieste sui filtri di dati.

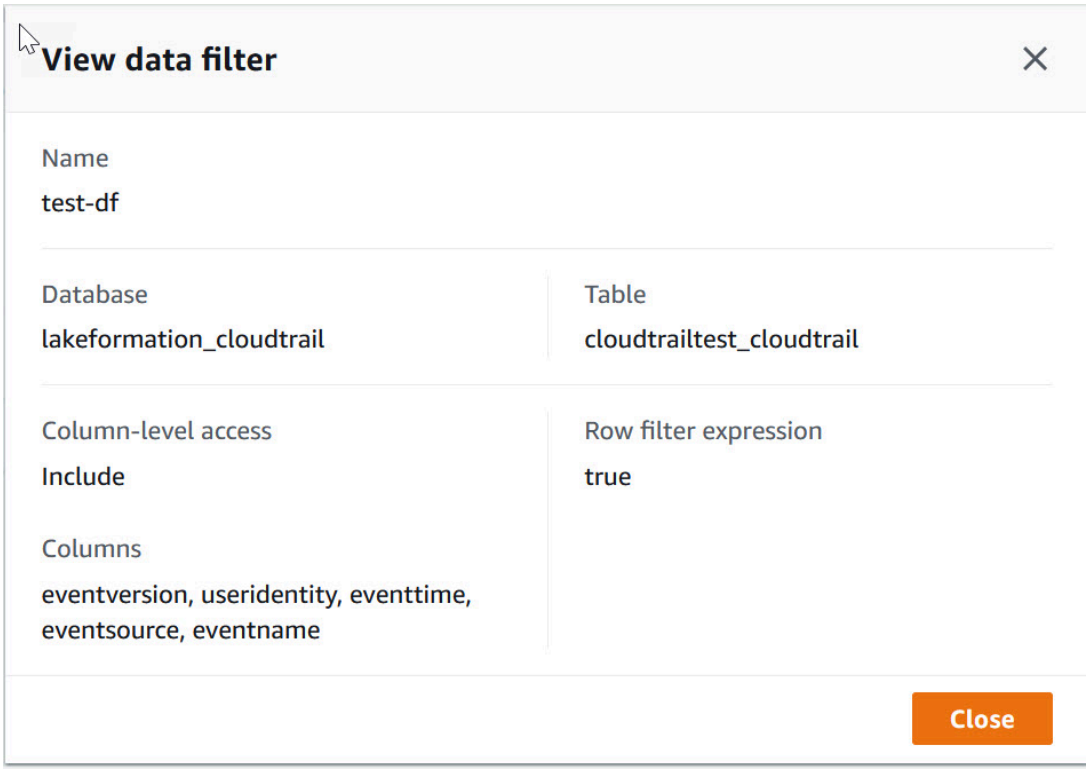
Console

1. Accedi AWS Management Console e apri la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.
2. Nel pannello di navigazione, in Catalogo dati, scegli Filtri dati.

   La pagina mostra i filtri di dati a cui hai accesso.



3. Per visualizzare i dettagli del filtro dati, scegli il filtro dati, quindi scegli Visualizza. Viene visualizzata una nuova finestra con informazioni dettagliate sul filtro dati.



AWS CLI

Immettete un `list-data-cells-filter` comando e specificate una risorsa della tabella.

L'esempio seguente elenca i filtri di dati per la `cloudtrailtest_cloudtrail` tabella.

```
aws lakeformation list-data-cells-filter --table '{ "CatalogId":"123456789012",
"DatabaseName":"lakeformation_cloudtrail", "Name":"cloudtrailtest_cloudtrail"}'
```

API/SDK

Utilizzate l'`ListDataCellsFilter`API e specificate una risorsa per la tabella.

L'esempio seguente usa Python per elencare i primi 20 filtri di dati per la `myTable` tabella.

```
response = client.list_data_cells_filter(
    Table = {
        'CatalogId': '111122223333',
        'DatabaseName': 'mydb',
        'Name': 'myTable'
    },
    MaxResults=20
```

```
)
```

Elenco delle autorizzazioni per il filtro dei dati

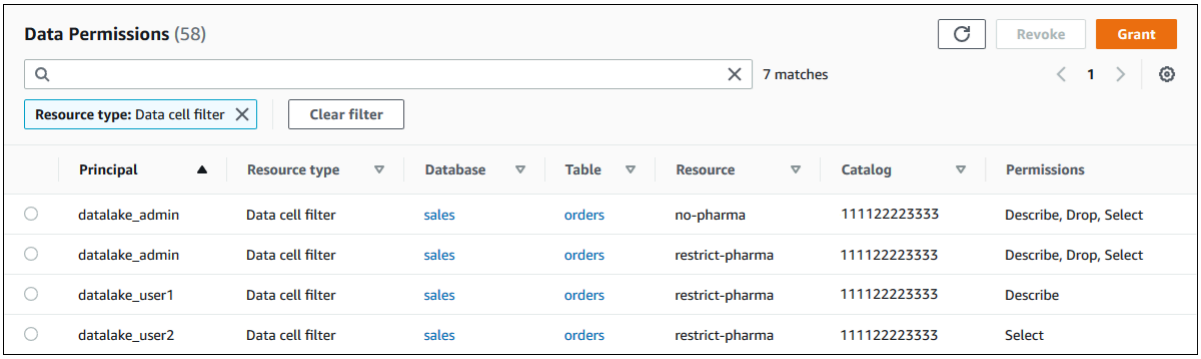
Puoi utilizzare la console Lake Formation per visualizzare le autorizzazioni concesse sui filtri di dati.

Per visualizzare le autorizzazioni su un filtro dati, devi essere un amministratore di Data Lake o disporre delle autorizzazioni richieste sul filtro dati.

Console

1. Accedi AWS Management Console e apri la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.
2. Nel pannello di navigazione, sotto Autorizzazioni, scegli Autorizzazioni dati.
3. Nella pagina Autorizzazioni dati, fai clic o tocca nel campo di ricerca e nel menu Proprietà scegli Tipo di risorsa.
4. Nel menu Tipo di risorsa, scegli Tipo di risorsa: filtro delle celle di dati.

   Sono elencati i filtri di dati per i quali disponi delle autorizzazioni. Potrebbe essere necessario scorrere orizzontalmente per visualizzare le colonne Autorizzazioni e Concedibili.



AWS CLI

- Immettete un comando. `list-permissions` Specificate `DataCellsFilter` per l'`resource`argomento e specificate `DESCRIBE` o `DROP` per l'`Permissions`argomento e, facoltativamente, per l'`PermissionsWithGrantOption`argomento.

  L'esempio seguente elenca `DESCRIBE` le autorizzazioni con l'opzione grant sul filtro dati. `restrict-pharma` I risultati sono limitati alle autorizzazioni concesse per il

principale `datalake_user1` e la `orders` tabella del `sales` database nell'AWSaccount 1111-2222-3333.

```
aws lakeformation list-permissions --cli-input-json file://list-params.json
```

Di seguito sono riportati i contenuti del file. `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"]
}
```

Visualizzazione delle autorizzazioni per database e tabelle in Lake Formation

È possibile visualizzare le autorizzazioni di Lake Formation concesse su un database o una tabella di Data Catalog. Puoi farlo utilizzando la console Lake Formation, l'API o AWS Command Line Interface (AWS CLI).

Utilizzando la console, è possibile visualizzare le autorizzazioni a partire dalle pagine Database o Tabelle o dalla pagina Autorizzazioni dati.

> **Note**
>
> Se non sei un amministratore di database o il proprietario della risorsa, puoi visualizzare le autorizzazioni che altri principali hanno sulla risorsa solo se disponi dell'autorizzazione Lake Formation sulla risorsa con l'opzione di concessione.

Oltre alle autorizzazioni Lake Formation richieste, sono necessarie le autorizzazioni AWS Identity and Access Management (IAM)`glue:GetDatabases`, `glue:GetDatabase` `glue:GetTablesglue:GetTable`, e. `glue:ListPermissions`

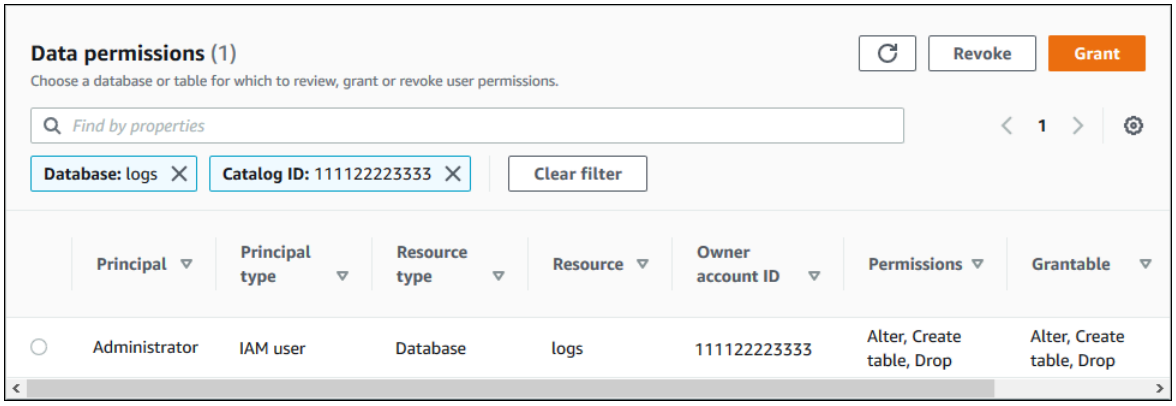
Per visualizzare le autorizzazioni su un database (console, a partire dalla pagina Database)

1. Aprire la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.

   Accedi come amministratore del data lake, creatore del database o come utente con un'autorizzazione Lake Formation sul database con l'opzione di concessione.
2. Nel riquadro di navigazione, scegli Databases (Database).
3. Scegli un database e nel menu Azioni scegli Visualizza autorizzazioni.

> **Note**
>
> Se scegli un link alla risorsa del database, Lake Formation visualizza le autorizzazioni sul link alla risorsa, non sul database di destinazione del link alla risorsa.

La pagina delle autorizzazioni dei dati elenca tutte le autorizzazioni di Lake Formation per il database. Il nome del database e l'ID del catalogo (ID AWS account) del proprietario del database vengono visualizzati come etichette sotto la casella di ricerca. I riquadri indicano che è stato applicato un filtro alle autorizzazioni di elenco solo per quel database. Puoi regolare il filtro chiudendo un riquadro o scegliendo Cancella filtro.



Per visualizzare le autorizzazioni su un database (console, a partire dalla pagina Autorizzazioni dati)

1. Aprire la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.

   Accedi come amministratore del data lake, creatore del database o come utente con un'autorizzazione Lake Formation sul database con l'opzione di concessione.
2. Nel pannello di navigazione, scegli Autorizzazioni dati.
3. Posiziona il cursore nella casella di ricerca nella parte superiore della pagina e nel menu Proprietà visualizzato, scegli Database.
4. Nel menu Database visualizzato, scegli un database.

   > **Note**
   >
   > Se scegli un link alla risorsa del database, Lake Formation visualizza le autorizzazioni sul link alla risorsa, non sul database di destinazione del link alla risorsa.

   La pagina delle autorizzazioni dei dati elenca tutte le autorizzazioni di Lake Formation per il database. Il nome del database viene visualizzato come riquadro sotto la casella di ricerca. Il riquadro indica che è stato applicato un filtro per elencare le autorizzazioni solo per quel database. Puoi rimuovere il filtro chiudendo il riquadro o scegliendo Cancella filtro.

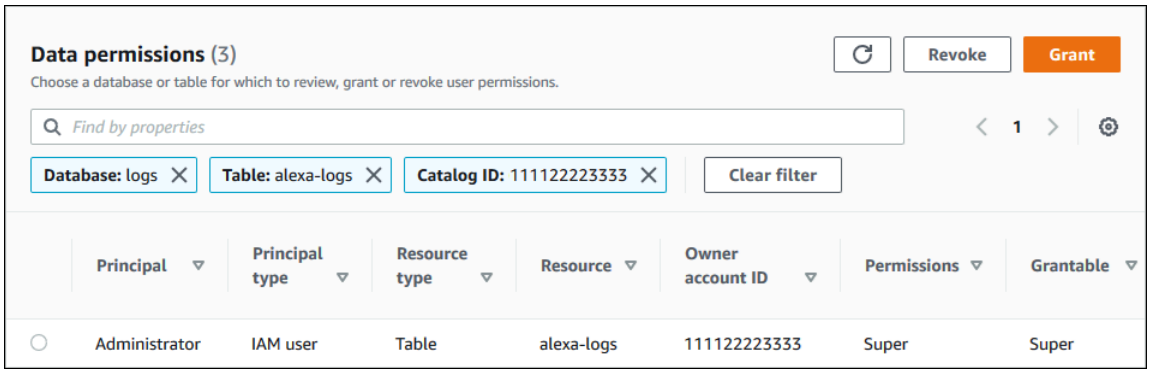
Per visualizzare le autorizzazioni su una tabella (console, a partire dalla pagina Tabelle)

1. Aprire la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.

   Accedi come amministratore del data lake, creatore della tabella o come utente con un'autorizzazione Lake Formation sulla tabella con l'opzione di concessione.
2. Nel pannello di navigazione, seleziona Tabelle.
3. Scegli una tabella e nel menu Azioni scegli Visualizza autorizzazioni.

   > **Note**
   >
   > Se scegli un link alla risorsa alla tabella, Lake Formation visualizza le autorizzazioni sul link alla risorsa, non sulla tabella di destinazione del link alla risorsa.

La pagina Autorizzazioni dati elenca tutte le autorizzazioni di Lake Formation per la tabella. Il nome della tabella, il nome del database che contiene la tabella e l'ID del catalogo (ID AWS account) del proprietario della tabella vengono visualizzati come etichette sotto la casella di ricerca. Le etichette indicano che è stato applicato un filtro alle autorizzazioni di elenco solo per quella tabella. Puoi regolare il filtro chiudendo un'etichetta o scegliendo Cancella filtro.



Per visualizzare le autorizzazioni su una tabella (console, a partire dalla pagina Autorizzazioni dati)

1. Aprire la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.

   Accedi come amministratore del data lake, creatore della tabella o come utente con un'autorizzazione Lake Formation sulla tabella con l'opzione di concessione.
2. Nel pannello di navigazione, scegli Autorizzazioni dati.
3. Posiziona il cursore nella casella di ricerca nella parte superiore della pagina e nel menu Proprietà visualizzato, scegli Database.
4. Nel menu Database visualizzato, scegli un database.

> **Important**
>
> Se desideri visualizzare le autorizzazioni su una tabella condivisa con il tuo AWS account da un account esterno, devi scegliere il database nell'account esterno che contiene la tabella, non un collegamento di risorsa al database.

La pagina delle autorizzazioni dei dati elenca tutte le autorizzazioni di Lake Formation per il database.

5. Posiziona nuovamente il cursore nella casella di ricerca e nel menu Proprietà visualizzato, scegli Tabella.
6. Nel menu Tabelle visualizzato, scegli una tabella.

   La pagina Autorizzazioni dati elenca tutte le autorizzazioni di Lake Formation per la tabella. Il nome della tabella e il nome del database che contiene la tabella vengono visualizzati come riquadri sotto la casella di ricerca. I riquadri indicano che è stato applicato un filtro alle autorizzazioni di elenco solo per quella tabella. Puoi regolare il filtro chiudendo un riquadro o scegliendo Cancella filtro.

Per visualizzare le autorizzazioni su una tabella () AWS CLI

- Immettere un `list-permissions` comando.

  L'esempio seguente elenca le autorizzazioni su una tabella condivisa da un account esterno. La `CatalogId` proprietà è l'AWSID dell'account esterno e il nome del database si riferisce al database dell'account esterno che contiene la tabella.

```
aws lakeformation list-permissions  --resource-type TABLE --resource '{ "Table":
 {"DatabaseName":"logs", "Name":"alexa-logs", "CatalogId":"123456789012"}}'
```

Revoca dell'autorizzazione utilizzando la console Lake Formation

Puoi utilizzare la console per revocare tutti i tipi di autorizzazioni di Lake Formation: autorizzazioni Data Catalog, autorizzazioni per policy tag, autorizzazioni per filtri dati e autorizzazioni relative alla posizione.

Per revocare le autorizzazioni di Lake Formation su una risorsa (console)

1. Aprire la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.

   Accedi come amministratore del data lake o come utente a cui sono state concesse le autorizzazioni con l'opzione di concessione sulla risorsa.
2. Nel pannello di navigazione, in Autorizzazioni, scegli Autorizzazioni Data lake, Tag e autorizzazioni LF o Posizioni dati.
3. Seleziona l'autorizzazione o la posizione, quindi scegli Revoca.
4. Nella finestra di dialogo che si apre, scegli Revoca.

Condivisione dei dati tra account in Lake Formation

Le funzionalità cross-account di Lake Formation consentono agli utenti di condividere in modo sicuro i data lake distribuiti tra più AWS organizzazioni o direttamente con i responsabili IAM in un altro account Account AWS, fornendo un accesso granulare ai metadati del Data Catalog e ai dati sottostanti. Le grandi aziende in genere utilizzano più Account AWS account e molti di questi account potrebbero aver bisogno di accedere a un data lake gestito da un singolo account. Account AWS Gli utenti e i job di AWS Glue estrazione, trasformazione e caricamento (ETL) possono eseguire query e unire tabelle su più account e sfruttare comunque le protezioni dei dati a livello di tabella e colonna di Lake Formation.

Quando concedi le autorizzazioni di Lake Formation su una risorsa Data Catalog a un account esterno o direttamente a un responsabile IAM in un altro account, Lake Formation utilizza il servizio AWS Resource Access Manager (AWS RAM) per condividere la risorsa. Se l'account del beneficiario appartiene alla stessa organizzazione dell'account concedente, la risorsa condivisa è immediatamente disponibile per il beneficiario. Se l'account del beneficiario non appartiene alla stessa organizzazione, AWS RAM invia un invito all'account del beneficiario per accettare o rifiutare la concessione di risorse. Quindi, per rendere disponibile la risorsa condivisa, l'amministratore del data lake nell'account del beneficiario deve utilizzare la console o accettare l' AWS RAM invito. AWS CLI

Lake Formation supporta la condivisione delle risorse del Data Catalog con account esterni in modalità di accesso ibrido. La modalità di accesso ibrido offre la flessibilità necessaria per abilitare selettivamente le autorizzazioni di Lake Formation per database e tabelle del tuo. AWS Glue Data Catalog
Con la modalità di accesso ibrida, ora disponi di un percorso incrementale che ti consente di impostare le autorizzazioni di Lake Formation per un set specifico di utenti senza interrompere le politiche di autorizzazione di altri utenti o carichi di lavoro esistenti.

Per ulteriori informazioni, consulta Modalità di accesso ibrida.

Condivisione diretta tra account

I responsabili autorizzati possono condividere le risorse in modo esplicito con un responsabile IAM in un account esterno. Questa funzionalità è utile quando il proprietario di un account desidera avere il controllo su chi nell'account esterno può accedere alle risorse. Le autorizzazioni ricevute dal preside IAM saranno costituite da un'unione di concessioni dirette e concessioni a livello di account, che verranno trasferite a cascata ai principali. L'amministratore del data lake dell'account del destinatario

può visualizzare le concessioni dirette tra account, ma non può revocare le autorizzazioni. Il principale che riceve la condivisione di risorse non può condividere la risorsa con altri destinatari.

Metodi per condividere le risorse del Data Catalog

Con un'unica operazione di concessione di Lake Formation, puoi concedere autorizzazioni tra account sulle seguenti risorse del Data Catalog.

- Un database
- Una tabella singola (con filtro opzionale per le colonne)
- Alcune tabelle selezionate
- Tutte le tabelle di un database (utilizzando il carattere jolly Tutte le tabelle)

Esistono due opzioni per condividere database e tabelle con un altro account Account AWS o con i principali IAM di un altro account.

- Controllo degli accessi basato su tag Lake Formation (LF-TBAC) (consigliato)

  Il controllo degli accessi basato su tag Lake Formation è una strategia di autorizzazione che definisce le autorizzazioni in base agli attributi. Puoi utilizzare il controllo degli accessi basato su tag per condividere le risorse del Data Catalog (database, tabelle e colonne) con responsabili IAM esterni, Account AWS Organizzazioni e unità organizzative (OU). In Lake Formation, questi attributi sono chiamati LF-tag. Per ulteriori informazioni, consulta Gestione di un data lake utilizzando il controllo degli accessi basato su tag Lake Formation.

  > **Note**
  >
  > Il metodo LF-TBAC per la concessione delle autorizzazioni di Data Catalog viene utilizzato per le concessioni tra account. AWS Resource Access Manager
  > Lake Formation ora supporta la concessione di autorizzazioni tra account a Organizzazioni e unità organizzative utilizzando il metodo LF-TBAC.
  > Per abilitare questa funzionalità, è necessario aggiornare le impostazioni della versione dell'account Cross alla versione 3.
  > Per ulteriori informazioni, consulta Aggiornamento delle impostazioni della versione di condivisione dei dati tra account.

- Risorse denominate Lake Formation

La condivisione dei dati tra account di Lake Formation utilizzando il metodo delle risorse denominate consente di concedere le autorizzazioni di Lake Formation con un'opzione di concessione su tabelle e database di Data Catalog a dirigenti Account AWS, organizzazioni o unità organizzative IAM esterni. L'operazione di concessione condivide automaticamente tali risorse.

> **Note**
>
> Puoi anche consentire al AWS Glue crawler di accedere a un data store in un account diverso utilizzando le credenziali di Lake Formation. Per ulteriori informazioni, consulta la sezione Scansione tra account nella Guida per gli sviluppatori. AWS Glue

I servizi integrati come Athena e Amazon Redshift Spectrum richiedono collegamenti alle risorse per poter includere risorse condivise nelle query. Per ulteriori informazioni sui link alle risorse, consulta. Come funzionano i link alle risorse in Lake Formation

Per considerazioni e limitazioni, vedereBuone pratiche e considerazioni sulla condivisione dei dati tra account.

Argomenti

- Prerequisiti
- Aggiornamento delle impostazioni della versione di condivisione dei dati tra account
- Condivisione delle tabelle e dei database del Data Catalog tra i nostri principali IAM provenienti Account AWS da account esterni
- Concessione delle autorizzazioni su un database o una tabella condivisa con il tuo account
- Concessione delle autorizzazioni per i collegamenti alle risorse
- Accesso ai dati sottostanti di una tabella condivisa
- Registrazione su più account CloudTrail
- Gestione delle autorizzazioni tra account utilizzando entrambi AWS Glue e Lake Formation
- Visualizzazione di tutte le sovvenzioni tra account utilizzando l'operazione API GetResourceShares

> **Argomenti correlati**
>
> - Panoramica delle autorizzazioni di Lake Formation

- Accesso e visualizzazione di tabelle e database condivisi del Data Catalog
- Creazione di collegamenti alle risorse
- Risoluzione dei problemi di accesso tra account

Prerequisiti

Prima che l' AWS account possa condividere le risorse di Data Catalog (database e tabelle) con un altro account o i responsabili di un altro account e prima di poter accedere alle risorse condivise con il proprio account, devono essere soddisfatti i seguenti prerequisiti.

Requisiti generali per la condivisione dei dati tra account

- Per condividere i database e le tabelle di Data Catalog in modalità di accesso ibrido, è necessario aggiornare le impostazioni della versione dell'account Cross alla versione 4.
- Prima di concedere autorizzazioni per più account su una risorsa Data Catalog, devi revocare tutte le autorizzazioni di Lake Formation dal `IAMAllowedPrincipals` gruppo per la risorsa. Se il principale chiamante dispone di autorizzazioni multiaccount per accedere a una risorsa e l'`IAMAllowedPrincipals`autorizzazione esiste sulla risorsa, Lake Formation lancia`AccessDeniedException`.

  Questo requisito è applicabile solo quando si registra la posizione dei dati sottostante in modalità Lake Formation. Se si registra la posizione dei dati in modalità ibrida, le autorizzazioni di `IAMAllowedPrincipals` gruppo possono esistere sul database o sulla tabella condivisi.
- Per i database che contengono tabelle che si intende condividere, è necessario impedire che alle nuove tabelle venga assegnato di default `Super` a`IAMAllowedPrincipals`. Sulla console Lake Formation, modifica il database e disattiva Usa solo il controllo di accesso IAM per le nuove tabelle in questo database o inserisci il seguente AWS CLI comando, sostituendolo `database` con il nome del database. Se la posizione dei dati sottostante è registrata in modalità di accesso ibrida, non è necessario modificare questa impostazione predefinita. In modalità di accesso ibrido, Lake Formation consente di applicare selettivamente le autorizzazioni di Lake Formation e le politiche di autorizzazione IAM per Amazon S3 e sulla stessa risorsa. AWS Glue

```
aws glue update-database --name database --database-input
 '{"Name":"database","CreateTableDefaultPermissions":[]}'
```

- Per concedere le autorizzazioni su più account, il concedente deve disporre delle autorizzazioni richieste (IAM) sul servizio. AWS Identity and Access Management AWS Glue AWS RAM La policy AWS `AWSLakeFormationCrossAccountManager` gestita concede le autorizzazioni richieste.

  Gli amministratori di Data Lake negli account che ricevono condivisioni di risorse utilizzando AWS RAM devono disporre della seguente politica aggiuntiva. Consente all'amministratore di accettare gli inviti alla condivisione AWS RAM delle risorse. Consente inoltre all'amministratore di abilitare la condivisione delle risorse con le organizzazioni.

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

- Se si desidera condividere le risorse di Data Catalog con le AWS Organizations nostre unità organizzative, è necessario abilitare la condivisione con le organizzazioni AWS RAM.

  Per informazioni su come abilitare la condivisione con le organizzazioni, consulta Abilitare la condivisione con AWS le organizzazioni nella Guida per l'AWS RAM utente.

  È necessario disporre dell'`ram:EnableSharingWithAwsOrganization`autorizzazione per abilitare la condivisione con le organizzazioni.

- Per condividere le risorse direttamente con un responsabile IAM in un altro account, devi aggiornare le impostazioni della versione dell'account Cross alla versione 3. Questa impostazione è disponibile nella pagina delle impostazioni del catalogo dati. Se utilizzi la versione 1, consulta le istruzioni per aggiornare l'impostazioneAggiornamento delle impostazioni della versione di condivisione dei dati tra account.
- Non è possibile condividere le risorse del Data Catalog crittografate con una chiave gestita dal AWS Glue servizio con un altro account. È possibile condividere solo le risorse del Data

Catalog crittografate con la chiave di crittografia del cliente e l'account che riceve la condivisione delle risorse deve disporre delle autorizzazioni sulla chiave di crittografia del Data Catalog per decrittografare gli oggetti.

Condivisione dei dati tra account utilizzando i requisiti LF-TBAC

- Per condividere le risorse di Data Catalog con AWS Organizations le unità organizzative (OU), è necessario aggiornare le impostazioni della versione dell'account Cross alla versione 3.
- Per condividere le risorse di Data Catalog con la versione 3 delle impostazioni della versione dell'account Cross, il concedente deve disporre delle autorizzazioni IAM definite `AWSLakeFormationCrossAccountManager` nella politica AWS gestita del proprio account.
- Se utilizzi la versione 1 o la versione 2 delle impostazioni della versione dell'account Cross, devi disporre di una politica delle risorse di Data Catalog (`glue:PutResourcePolicy`) che abiliti LF-TBAC. Per ulteriori informazioni, consulta Gestione delle autorizzazioni tra account utilizzando entrambi AWS Glue e Lake Formation.
- Se attualmente utilizzi una politica delle risorse di AWS Glue Data Catalog per condividere risorse e desideri concedere autorizzazioni tra account utilizzando la versione 3 delle impostazioni della versione Cross account, devi aggiungere l'**`glue:ShareResource`**autorizzazione nelle Impostazioni del Data Catalog utilizzando l'operazione `glue:PutResourcePolicy` API, come mostrato nella sezione. Gestione delle autorizzazioni tra account utilizzando entrambi AWS Glue e Lake Formation Questa politica non è richiesta se il tuo account non ha concesso concessioni tra account utilizzando la politica delle risorse di AWS Glue Data Catalog (`glue:PutResourcePolicy`autorizzazione all'uso delle versioni 1 e 2) per concedere l'accesso a più account.

```
{
      "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 il tuo account ha effettuato condivisioni tra account utilizzando la politica delle risorse di AWS Glue Data Catalog e attualmente utilizzi il metodo Named Resource o LF-TBAC con impostazioni Cross account versione 3 per condividere le risorse, che utilizza AWS RAM per condividere le risorse, devi impostare l'`EnableHybrid`argomento su quando richiami l'operazione API. `'true'` `glue:PutResourcePolicy` Per ulteriori informazioni, consulta Gestione delle autorizzazioni tra account utilizzando entrambi AWS Glue e Lake Formation.

Configurazione richiesta in ogni account che accede alla risorsa condivisa

- Se condividi risorse con Account AWS, almeno un utente dell'account consumer deve essere un amministratore del data lake per visualizzare le risorse condivise. Per informazioni su come creare un amministratore di data lake, consultaCrea un amministratore del data lake.

  L'amministratore del data lake può concedere le autorizzazioni di Lake Formation sulle risorse condivise ad altri responsabili dell'account. Gli altri responsabili non possono accedere alle risorse condivise finché l'amministratore del data lake non concede loro le autorizzazioni sulle risorse.
- I servizi integrati come Athena e Redshift Spectrum richiedono collegamenti alle risorse per poter includere risorse condivise nelle query. I responsabili devono creare un collegamento di risorsa nel proprio catalogo dati a una risorsa condivisa da un'altra risorsa. Account AWS Per ulteriori informazioni sui collegamenti alle risorse, consultaCome funzionano i link alle risorse in Lake Formation.
- Quando una risorsa viene condivisa direttamente con un principale IAM, per interrogare la tabella utilizzando Athena, il principale deve creare un collegamento alla risorsa. Per creare un collegamento a una risorsa, il principale necessita dell'`CREATE_DATABASE`autorizzazione `CREATE_TABLE` o del Lake Formation e dell'autorizzazione `glue:CreateTable` o `glue:CreateDatabase` IAM.

  Se l'account produttore condivide una tabella diversa dello stesso database con lo stesso principale o con un altro principale, tale principale può interrogare immediatamente la tabella.

> **Note**
>
> Per l'amministratore del data lake e per i responsabili a cui l'amministratore del data lake ha concesso le autorizzazioni, le risorse condivise vengono visualizzate nel Data Catalog come

se fossero risorse locali (di proprietà). I job di estrazione, trasformazione e caricamento (ETL) possono accedere ai dati sottostanti delle risorse condivise.
Per le risorse condivise, le pagine Tabelle e database della console Lake Formation visualizzano l'ID dell'account del proprietario.
Quando si accede ai dati sottostanti di una risorsa condivisa, gli eventi di CloudTrail registro vengono generati sia nell'account del destinatario della risorsa condivisa che nell'account del proprietario della risorsa. Gli CloudTrail eventi possono contenere l'ARN del principale che ha avuto accesso ai dati, ma solo se l'account del destinatario sceglie di includere l'ARN principale nei registri. Per ulteriori informazioni, consulta Registrazione su più account CloudTrail .

Aggiornamento delle impostazioni della versione di condivisione dei dati tra account

Di tanto in tanto, AWS Lake Formation aggiorna le impostazioni di condivisione dei dati tra account per distinguere le modifiche apportate all' AWS RAM utilizzo e per supportare gli aggiornamenti apportati alla funzionalità di condivisione dei dati tra account. Quando Lake Formation esegue questa operazione, crea una nuova versione delle impostazioni della versione dell'account Cross.

Principali differenze tra le impostazioni delle versioni per più account

Per ulteriori informazioni su come funziona la condivisione dei dati tra account in diverse impostazioni di versione di Cross account, consulta le seguenti sezioni.

> **Note**
>
> Per condividere i dati con un altro account, il concedente deve avere `AWSLakeFormationCrossAccountManager` gestito le autorizzazioni relative alla policy IAM. Questo è un prerequisito per tutte le versioni.
> L'aggiornamento delle impostazioni della versione dell'account Cross non influisce sulle autorizzazioni del destinatario sulle risorse condivise. Ciò è applicabile quando si esegue l'aggiornamento dalla versione 1 alla versione 2, dalla versione 2 alla versione 3 e dalla versione 1 alla versione 3. Per l'aggiornamento delle versioni, consulta le considerazioni elencate di seguito.

Versione 1

Metodo di risorsa denominato: associa ogni autorizzazione concessa tra più account di Lake Formation a una condivisione di AWS RAM risorse. L'utente (concedente, ruolo o responsabile) non richiede autorizzazioni aggiuntive.

Metodo LF-TBAC: le concessioni di autorizzazioni di Lake Formation tra account non vengono utilizzate per condividere dati. AWS RAM L'`glue:PutResourcePolicy`utente deve avere l'autorizzazione.

Vantaggi dell'aggiornamento delle versioni: versione iniziale, non applicabile.

Considerazioni sull'aggiornamento delle versioni: Versione iniziale - non applicabile

Versione 2

Metodo della risorsa denominata: ottimizza il numero di condivisioni di AWS RAM risorse mappando più concessioni di autorizzazioni tra account con un'unica condivisione di risorse. AWS RAM L'utente non richiede autorizzazioni aggiuntive.

Metodo LF-TBAC: le concessioni di autorizzazioni di Lake Formation tra account non vengono utilizzate per condividere dati. AWS RAM L'`glue:PutResourcePolicy`utente deve avere l'autorizzazione.

Vantaggi dell'aggiornamento delle versioni: configurazione scalabile tra più account grazie all'utilizzo ottimale della capacità. AWS RAM

Considerazioni sull'aggiornamento delle versioni: gli utenti che desiderano concedere le autorizzazioni Lake Formation su più account devono disporre delle autorizzazioni nella politica gestita. `AWSLakeFormationCrossAccountManager` AWS Altrimenti, sono necessari i `ram:DisassociateResourceShare` permessi necessari per condividere correttamente `ram:AssociateResourceShare` le risorse con un altro account.

Versione 3

Metodo della risorsa denominata: ottimizza il numero di condivisioni di AWS RAM risorse mappando più concessioni di autorizzazioni tra account con un'unica condivisione di risorse. AWS RAM L'utente non richiede autorizzazioni aggiuntive.

Metodo LF-TBAC: Lake Formation utilizza AWS RAM per le sovvenzioni tra account. L'utente deve aggiungere la dichiarazione glue: all'autorizzazione. ShareResource

`glue:PutResourcePolicy` Il destinatario deve accettare gli inviti alla condivisione delle risorse da AWS RAM.

Vantaggi dell'aggiornamento delle versioni: supporta le seguenti funzionalità:

- Consente di condividere le risorse in modo esplicito con un principale IAM in un account esterno.

  Per ulteriori informazioni, consulta Concessione e revoca delle autorizzazioni per le risorse del Data Catalog.
- Abilita le condivisioni tra account utilizzando il metodo LF-TBAC per Organizzazioni o unità organizzative (OU).
- Elimina il sovraccarico derivante dal mantenimento di politiche aggiuntive per le sovvenzioni tra account. AWS Glue

Considerazioni sull'aggiornamento delle versioni: se il concedente utilizza una versione precedente alla versione 3 e il destinatario utilizza la versione 3 o successiva, il concedente riceve il seguente messaggio di errore: «Richiesta di concessione tra account non valida. Per l'account consumer è disponibile la versione cross-account: v3. Effettua l'aggiornamento `CrossAccountVersion` `DataLakeSetting` alla versione minima v3 (Servizio: AmazonDataCatalog; Codice di stato: 400; Codice di errore: InvalidInputException)». Tuttavia, se il concedente utilizza la versione 3 e il destinatario utilizza la versione 1 o la versione 2, le sovvenzioni tra account vengono approvate correttamente.

Per condividere le risorse direttamente con i responsabili IAM di un altro account, solo il concedente deve utilizzare la versione 3.

Le sovvenzioni tra account effettuate utilizzando il metodo LF-TBAC richiedono agli utenti di disporre di una politica delle risorse nell'account. AWS Glue Data Catalog Quando si esegue l'aggiornamento alla versione 3, LF-TBAC concede gli utilizzi. AWS RAM Per consentire l'esecuzione delle sovvenzioni AWS RAM basate su più account, è necessario aggiungere la `glue:ShareResource` dichiarazione alle politiche esistenti in materia di risorse del Data Catalog, come mostrato nella sezione. Gestione delle autorizzazioni tra account utilizzando entrambi AWS Glue e Lake Formation

Versione 4

Il concedente necessita della versione 4 o successiva per condividere le risorse di Data Catalog in modalità di accesso ibrido.

Ottimizza la condivisione delle risorse AWS RAM

Le nuove versioni (versione 2 e successive) delle sovvenzioni tra account utilizzano in modo ottimale la AWS RAM capacità per massimizzare l'utilizzo tra account. Quando condividi una risorsa con un responsabile IAM Account AWS o esterno, Lake Formation può creare una nuova condivisione di risorse o associare la risorsa a una condivisione esistente. Associandosi alle azioni esistenti, Lake Formation riduce il numero di inviti alla condivisione delle risorse che un consumatore deve accettare.

Abilita AWS RAM le condivisioni tramite TBAC o condividi le risorse direttamente con i principali

Per condividere le risorse direttamente con i responsabili IAM in un altro account o per abilitare le condivisioni TBAC tra account su Organizations o unità organizzative, devi aggiornare le impostazioni della versione Cross account alla versione 3. Per ulteriori informazioni sui limiti delle AWS RAM risorse, consulta. Buone pratiche e considerazioni sulla condivisione dei dati tra account

Autorizzazioni necessarie per l'aggiornamento delle impostazioni delle versioni tra account

Se un concedente di autorizzazioni su più account ha `AWSLakeFormationCrossAccountManager` gestito le autorizzazioni relative alle policy IAM, non è richiesta alcuna configurazione aggiuntiva delle autorizzazioni per il ruolo o il responsabile del concedente delle autorizzazioni su più account. Tuttavia, se il concedente che concede più account non utilizza la policy gestita, affinché la nuova versione della concessione tra più account abbia successo, al ruolo o al responsabile del concedente devono essere concesse le seguenti autorizzazioni IAM.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "VisualEditor1",
      "Effect": "Allow",
      "Action": [
         "ram:AssociateResourceShare",
         "ram:DisassociateResourceShare",
         "ram:GetResourceShares"
       ],
     "Resource": "*",
     "Condition": {
       "StringLike": {
```

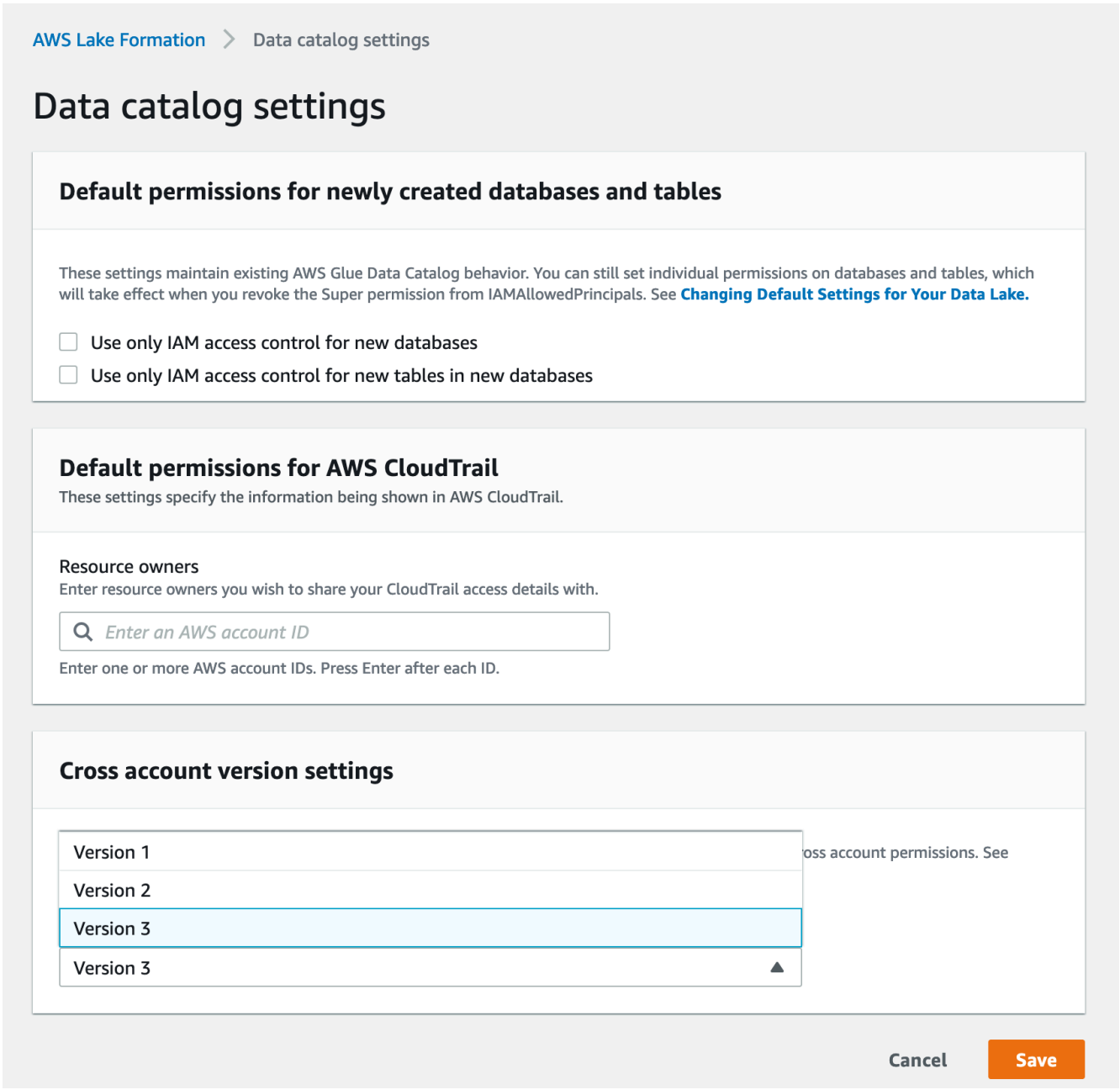
```
          "ram:ResourceShareName": "LakeFormation*"
        }
      }
    }
  ]
}
```

Per abilitare la nuova versione

Segui questi passaggi per aggiornare le impostazioni della versione dell'account Cross tramite la AWS Lake Formation console o il AWS CLI.

Console

1. Scegli la versione 2, la versione 3 o la versione 4 nelle impostazioni della versione dell'account Cross nella pagina delle impostazioni del catalogo dati. Se selezioni la versione 1, Lake Formation utilizzerà la modalità di condivisione delle risorse predefinita.



2. Selezionare Salva.

AWS Command Line Interface (AWS CLI)

Utilizzate il `put-data-lake-settings` AWS CLI comando per impostare il `CROSS_ACCOUNT_VERSION` parametro. I valori accettati sono 1, 2, 3 e 4.

```
aws lakeformation put-data-lake-settings --region us-east-1 --data-lake-settings
 file://settings
{
```

```
    "DataLakeAdmins": [
        {
            "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/test"
        }
    ],
    "CreateDatabaseDefaultPermissions": [],
    "CreateTableDefaultPermissions": [],
    "Parameters": {
        "CROSS_ACCOUNT_VERSION": "3"
    }
}
```

> **Important**
>
> Dopo aver scelto la versione 2 o la versione 3, tutte le nuove sovvenzioni per risorse denominate passeranno attraverso la nuova modalità di concessione tra account. Per utilizzare in modo ottimale la AWS RAM capacità delle condivisioni esistenti su più account, ti consigliamo di revocare le sovvenzioni concesse con la versione precedente e di riassegnarle nella nuova modalità.

Condivisione delle tabelle e dei database del Data Catalog tra i nostri principali IAM provenienti Account AWS da account esterni

Questa sezione include istruzioni su come abilitare le autorizzazioni tra account su tabelle e database di Data Catalog a un AWS account esterno, un responsabile IAM, un'organizzazione o un'unità organizzativa. L'operazione di concessione condivide automaticamente tali risorse.

Argomenti

- Condivisione dei dati tramite controllo degli accessi basato su tag
- Condivisione dei dati tra account utilizzando il metodo della risorsa denominato

Condivisione dei dati tramite controllo degli accessi basato su tag

Configurazione obbligatoria sull'account produttore/concedente

1. Definire un tag LF. Per istruzioni su come creare un tag LF, vedere. Creazione di tag LF

2. Assegnate il tag LF alla risorsa di destinazione. Per ulteriori informazioni, consulta Assegnazione di tag LF alle risorse del Data Catalog.
3. Concedi l'autorizzazione LF-Tag all'account esterno. Per ulteriori informazioni, consulta Concessione delle autorizzazioni LF-Tag tramite la console.

   A questo punto, l'amministratore del consumer data lake dovrebbe essere in grado di trovare il policy tag condiviso tramite la console dell'account beneficiario Lake Formation, in Autorizzazioni, ruoli e attività amministrativi, LF-Tags.
4. Concedi l'autorizzazione ai dati all'account esterno/beneficiario.
   a. Nel pannello di navigazione, in Autorizzazioni, Autorizzazioni Data lake, scegli Concedi.
   b. Per Principal, scegli Account esterni e inserisci l' Account AWS ID di destinazione o il ruolo IAM del principale o Amazon Resource Name (ARN) per il principale (ARN principale).
   c. Per i tag LF o le risorse del catalogo, scegli la chiave e i valori del tag LF da condividere con l'account consumatore (chiave e valore). **`Confidentiality`** `public`
   d. Per Autorizzazioni, in Risorse abbinate ai tag LF (consigliato) scegli Aggiungi tag LF.
   e. Seleziona la chiave e il valore del tag che viene condiviso con l'account del beneficiario (chiave e valore). `Confidentiality` `public`
   f. Per le autorizzazioni del database, seleziona Descrivi in Autorizzazioni del database per concedere le autorizzazioni di accesso a livello di database.
   g. L'amministratore del consumer data lake dovrebbe essere in grado di trovare il policy tag condiviso tramite l'account consumer sulla console di Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/, in Autorizzazioni, ruoli e attività amministrativi, LF-Tags.
   h. Seleziona Descrivi in Autorizzazioni concesse in modo che l'account consumer possa concedere autorizzazioni a livello di database ai suoi utenti.

      Poiché l'amministratore del data lake deve concedere le autorizzazioni sulle risorse condivise ai responsabili dell'account beneficiario, le autorizzazioni per più account devono sempre essere concesse con l'opzione di concessione.

      > **Note**
      >
      > I mandanti che ricevono sovvenzioni dirette tra più account non avranno l'opzione di autorizzazione Grantable.

i. Per le autorizzazioni per tabelle e colonne, seleziona Seleziona e descrivi in Autorizzazioni per le tabelle.

j. Seleziona Seleziona e descrivi in Autorizzazioni concedibili.

k. Scegli Concessione.

Configurazione obbligatoria sull'account destinatario/beneficiario

1. Quando condividi una risorsa con un altro account, la risorsa appartiene ancora all'account del produttore e non è visibile nella console Athena. Per rendere visibile la risorsa nella console Athena, devi creare un link alla risorsa che punti alla risorsa condivisa. Per istruzioni sulla creazione di un collegamento a una risorsa, consulta e Creazione di un collegamento di risorsa a una tabella condivisa del catalogo dati Creazione di un collegamento di risorsa a un database Data Catalog condiviso
2. È necessario creare un set separato di tag LF nell'account utente per utilizzare il controllo degli accessi basato su tag LF quando si condividono i link alle risorse. Crea e assegna i tag LF richiesti al database/alle tabelle condivisi e ai link alle risorse.
3. Concedi le autorizzazioni su questi LF-tag ai principali IAM nell'account del beneficiario.

Condivisione dei dati tra account utilizzando il metodo della risorsa denominato

È possibile concedere le autorizzazioni direttamente ai responsabili di un altro AWS account o a utenti esterni o. Account AWS AWS Organizations Concedere i permessi di Lake Formation a Organizzazioni o unità organizzative equivale a concedere il permesso Account AWS a tutti i membri di quell'organizzazione o unità organizzativa.

Quando concedi autorizzazioni ad account o organizzazioni esterne, devi includere l'opzione Autorizzazioni concedibili. Solo l'amministratore del data lake dell'account esterno può accedere alle risorse condivise finché l'amministratore non concede le autorizzazioni sulle risorse condivise ad altri responsabili dell'account esterno.

> **Note**
>
> L'opzione delle autorizzazioni concedibili non è supportata quando si concedono le autorizzazioni direttamente ai principali IAM da account esterni.

Segui le istruzioni per concedere le autorizzazioni Concessione delle autorizzazioni al database utilizzando il metodo di risorsa denominato per più account utilizzando il metodo di risorsa denominato.

Concessione delle autorizzazioni su un database o una tabella condivisa con il tuo account

Dopo che una risorsa Data Catalog appartenente a un altro AWS account è stata condivisa con il tuo AWS account, in qualità di amministratore del data lake, puoi concedere le autorizzazioni sulla risorsa condivisa ad altri responsabili del tuo account. Tuttavia, non puoi concedere le autorizzazioni sulla risorsa ad altri AWS account o organizzazioni.

Puoi utilizzare la AWS Lake Formation console, l'API o AWS Command Line Interface (AWS CLI) per concedere le autorizzazioni.

Per concedere le autorizzazioni su un database condiviso (denominato metodo di risorsa, console)

- Segui le istruzioni in Concessione delle autorizzazioni al database utilizzando il metodo di risorsa denominato. Nell'elenco dei database sotto i tag LF o le risorse del catalogo, assicuratevi di selezionare il database nell'account esterno, non un collegamento alla risorsa per il database.

  Se non vedi il database nell'elenco dei database, assicurati di aver accettato l'invito AWS Resource Access Manager (AWS RAM) alla condivisione delle risorse per il database. Per ulteriori informazioni, consulta Accettazione di un invito alla condivisione delle risorse daAWS RAM.

  Inoltre, per le `ALTER` autorizzazioni `CREATE_TABLE` e, segui le istruzioni riportate in Concessione delle autorizzazioni per la localizzazione dei dati (stesso account) e assicurati di inserire l'ID dell'account proprietario nel campo Ubicazione dell'account registrato.

Per concedere le autorizzazioni su una tabella condivisa (metodo di risorsa denominato, console)

- Segui le istruzioni in Concessione delle autorizzazioni per le tabelle utilizzando il metodo di risorsa denominato. Nell'elenco dei database sotto i tag LF o le risorse del catalogo, assicuratevi di selezionare il database nell'account esterno, non un collegamento alla risorsa per il database.

  Se non vedi la tabella nell'elenco delle tabelle, assicurati di aver accettato l'invito alla condivisione delle AWS RAM risorse per la tabella. Per ulteriori informazioni, consulta Accettazione di un invito alla condivisione delle risorse daAWS RAM.

Inoltre, per l'`ALTER`autorizzazione, segui le istruzioni riportate e assicurati di inserire l'ID dell'account proprietario nel campo Ubicazione dell'account registrato. Concessione delle autorizzazioni per la localizzazione dei dati (stesso account)

Per concedere autorizzazioni su risorse condivise (metodo LF-TBAC, console)

- Segui le istruzioni in Concessione delle autorizzazioni di Data Catalog . Nella sezione LF-Tag o risorse del catalogo, concedi l'espressione esatta del tag LF che l'account esterno ha concesso al tuo account o un sottoinsieme di quell'espressione.

  Ad esempio, se un account esterno ha concesso l'espressione LF-Tag `module=customers AND environment=production` al vostro account con l'opzione di concessione, in qualità di amministratore del data lake, potete concedere la stessa espressione `module=customers` o `environment=production` a un responsabile del vostro account. È possibile concedere solo le stesse autorizzazioni o un sottoinsieme delle autorizzazioni di Lake Formation (ad esempio, `SELECT`ALTER, e così via) concesse alle risorse tramite l'espressione LF-Tag.

Per concedere le autorizzazioni su una tabella condivisa (metodo denominato resource,) AWS CLI

- Utilizzare un comando simile al seguente: In questo esempio:
  - L'ID del tuo AWS account è 1111-2222-3333.
  - L'account proprietario della tabella e che l'ha concessa al tuo account è 1234-5678-9012.
  - L'`SELECT`autorizzazione sulla tabella condivisa viene concessa all'utente. `pageviews` `datalake_user1` Quell'utente è il principale del tuo account.
  - La `pageviews` tabella si trova nel `analytics` database, di proprietà dell'account 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"}}'
```

  Si noti che l'account proprietario deve essere specificato nella proprietà dell'argomento. `CatalogId` `resource`

Concessione delle autorizzazioni per i collegamenti alle risorse

Segui questi passaggi per concedere AWS Lake Formation le autorizzazioni su uno o più link di risorse a un responsabile del tuo account. AWS

Dopo aver creato un link a una risorsa, solo tu puoi visualizzarlo e accedervi. (Ciò presuppone che il controllo di accesso Use only IAM per le nuove tabelle in questo database non sia abilitato per il database.) Per consentire agli altri responsabili del tuo account di accedere al link alla risorsa, concedi almeno l'`DESCRIBE`autorizzazione.

> **Important**
>
> La concessione delle autorizzazioni su un collegamento a una risorsa non concede le autorizzazioni sulla tabella o sul database di destinazione (collegato). È necessario concedere le autorizzazioni sulla destinazione separatamente.

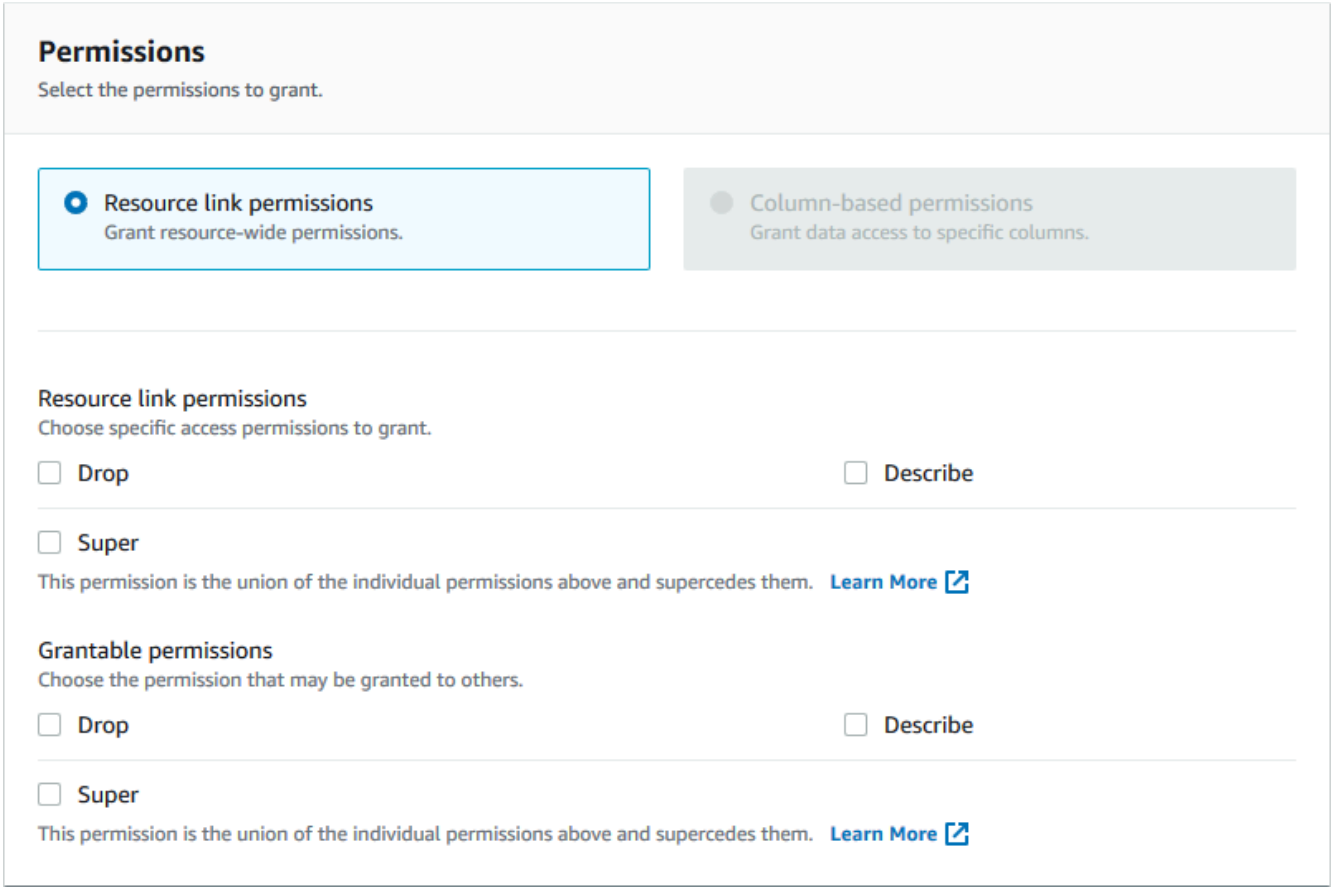
Puoi concedere le autorizzazioni utilizzando la console Lake Formation, l'API o AWS Command Line Interface (AWS CLI).

console

Per concedere le autorizzazioni per i collegamenti alle risorse utilizzando la console Lake Formation

1. Esegui una di queste operazioni:
   - Per i collegamenti alle risorse del database, segui la procedura riportata in Concessione delle autorizzazioni al database utilizzando il metodo di risorsa denominato. Per effettuare le seguenti operazioni:
     1. Apri la pagina Concedi le autorizzazioni del data lake.
     2. Specificare i database. Specificare uno o più collegamenti alle risorse del database.
     3. Specificare i principali.
   - Per i collegamenti alle risorse delle tabelle, Concessione delle autorizzazioni per le tabelle utilizzando il metodo di risorsa denominato procedi nel seguente modo:
     1. Apri la pagina Concedi le autorizzazioni del data lake.
     2. Specificare le tabelle. Specificate uno o più collegamenti alle risorse della tabella.
     3. Specificare i principali.

2. In Autorizzazioni, seleziona le autorizzazioni da concedere. Facoltativamente, seleziona autorizzazioni concedibili.



3. Scegli Concessione.

AWS CLI

Per concedere le autorizzazioni per i collegamenti alle risorse utilizzando AWS CLI

- Esegui il `grant-permissions` comando, specificando un link alla risorsa come risorsa.

  Example

  Questo esempio concede `DESCRIBE` all'utente `datalake_user1` sulla tabella il collegamento alla risorsa nel database `incidents-link` `issues` nell' AWS account 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"}}'
```

Consulta anche:

- Creazione di collegamenti alle risorse
- Riferimento alle autorizzazioni di Lake Formation

Accesso ai dati sottostanti di una tabella condivisa

Supponiamo che l' AWS account A condivida una tabella del catalogo dati con l'account B, ad esempio concedendo l'opzione di concessione `SELECT` sulla tabella all'account B. Affinché un responsabile dell'account B sia in grado di leggere i dati sottostanti della tabella condivisa, devono essere soddisfatte le seguenti condizioni:

- L'amministratore del data lake dell'account B deve accettare la condivisione. (Questo non è necessario se gli account A e B fanno parte della stessa organizzazione o se la concessione è stata concessa con il metodo di controllo degli accessi basato su tag Lake Formation.)
- L'amministratore del data lake deve concedere nuovamente al principale l'`SELECT`autorizzazione Lake Formation concessa dall'account A sulla tabella condivisa.
- Il principale deve disporre delle seguenti autorizzazioni IAM sulla tabella, sul database che la contiene e sull'account A Data Catalog.

> **Note**
>
> Nella seguente politica IAM:
>
> - Sostituisci <account-id-A>con l' AWS ID dell'account A.
> - Sostituisci <region>con una regione valida.
> - Sostituisci <database>con il nome del database nell'account A che contiene la tabella condivisa.
> - Sostituisci <table>con il nome della tabella condivisa.

```
{
```

```
    "Version": "2012-10-17",
    "Statement": [
        {
          "Effect": "Allow",
          "Action": [
            "glue:GetTable",
            "glue:GetTables",
            "glue:GetPartition",
            "glue:GetPartitions",
            "glue:BatchGetPartition",
            "glue:GetDatabase",
            "glue:GetDatabases"
           ],
           "Resource": [
            "arn:aws:glue:<region>:<account-id-A>:table/<database>/<table>",
            "arn:aws:glue:<region>:<account-id-A>:database/<database>",
            "arn:aws:glue:<region>:<account-id-A>:catalog"
           ]
        },
        {
          "Effect": "Allow",
          "Action": [
            "lakeformation:GetDataAccess"
           ],
          "Resource": [
            "*"
           ],
          "Condition": {
            "StringEquals": {
              "lakeformation:GlueARN":"arn:aws:glue:<region>:<account-id-
A>:table/<database>/<table>"
            }
        }
    }
   ]
}
```

Consulta anche:

- Accettazione di un invito alla condivisione delle risorse daAWS RAM

Registrazione su più account CloudTrail

Lake Formation fornisce un audit trail centralizzato di tutti gli accessi tra account ai dati nel tuo data lake. Quando un AWS account destinatario accede ai dati in una tabella condivisa, Lake Formation copia l' CloudTrail evento nei log dell' CloudTrail account proprietario. Gli eventi copiati includono interrogazioni sui dati da parte di servizi integrati come Amazon Redshift Spectrum Amazon Athena e accessi ai dati tramite processi. AWS Glue

CloudTrail gli eventi per le operazioni tra account sulle risorse di Data Catalog vengono copiati in modo analogo.

In qualità di proprietario delle risorse, se abiliti la registrazione a livello di oggetto in Amazon S3, puoi eseguire query che uniscono gli eventi S3 agli eventi di Lake CloudTrail Formation per determinare CloudTrail gli account che hanno avuto accesso ai tuoi bucket S3.

Argomenti

Inclusione delle identità principali nei log di più account CloudTrail

Per impostazione predefinita, gli CloudTrail eventi tra account aggiunti ai log del destinatario della risorsa condivisa e copiati nei log del proprietario della risorsa contengono solo l'ID AWS principale dell'account esterno, non il nome Amazon Resource Name (ARN) leggibile dall'uomo del principale (ARN principale). Quando condividi risorse all'interno di confini affidabili, ad esempio all'interno della stessa organizzazione o team, puoi scegliere di includere l'ARN principale negli CloudTrail eventi. Gli account dei proprietari delle risorse possono quindi tenere traccia dei principali account dei destinatari che accedono alle risorse di proprietà.

Important

In qualità di destinatario di risorse condivise, per visualizzare l'ARN principale negli eventi nei propri CloudTrail registri, è necessario scegliere di condividere l'ARN principale con l'account del proprietario.
Se l'accesso ai dati avviene tramite un collegamento alle risorse, nell'account del destinatario della risorsa condivisa vengono registrati due eventi: uno per l'accesso al collegamento alle risorse e uno per l'accesso alla risorsa di destinazione. L'evento per l'accesso al link di risorsa include l'ARN principale. L'evento per l'accesso alle risorse di destinazione non include l'ARN

principale senza l'opt-in. L'evento di accesso al collegamento alle risorse non viene copiato nell'account del proprietario.

Di seguito è riportato un estratto di un CloudTrail evento predefinito tra più account (senza opt-in). L'account che esegue l'accesso ai dati è 1111-2222-3333. Questo è il registro visualizzato sia nell'account chiamante che nell'account del proprietario della risorsa. Lake Formation compila i log in entrambi gli account nel caso di più account.

```
{
    "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"
    },
...
}
```

In qualità di consumatore di risorse condivise, quando scegli di includere l'ARN principale, l'estratto diventa il seguente. Il `lakeFormationPrincipal` campo rappresenta il ruolo finale o l'utente che esegue la query tramite Amazon Athena, Amazon Redshift Spectrum o job. 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"
    },
...
}
```

Per attivare l'inclusione degli ARN principali nei log di più account CloudTrail

1. Aprire la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.

   Accedi come utente o come `Administrator` utente con la policy IAM. `Administrator Access`
2. Nel pannello di navigazione scegli Impostazioni.
3. Nella pagina delle impostazioni del catalogo dati, nella AWS CloudTrail sezione Autorizzazioni predefinite per, per Proprietari delle risorse, inserisci uno o più ID di account del proprietario della AWS risorsa.

   Premi Invio dopo ogni ID account.
4. Selezionare Salva.

   Ora CloudTrail gli eventi tra account archiviati nei log sia per il destinatario della risorsa condivisa che per il proprietario della risorsa contengono l'ARN principale.

Interrogazione dei CloudTrail log per l'accesso a più account Amazon S3

In qualità di proprietario di risorse condivise, puoi interrogare CloudTrail i log di S3 per determinare gli account che hanno avuto accesso ai tuoi bucket Amazon S3 (a condizione che tu abbia abilitato la registrazione a livello di oggetto in Amazon S3). Questo vale solo per le sedi S3 che hai registrato con Lake Formation. Se i consumatori di risorse condivise scelgono di includere i Ran principali nei CloudTrail log di Lake Formation, puoi determinare i ruoli o gli utenti che hanno avuto accesso ai bucket.

Quando esegui query con Amazon Athena, puoi unire gli eventi Lake Formation e CloudTrail gli eventi S3 nella CloudTrail proprietà del nome della sessione. Le query possono anche filtrare gli eventi di Lake Formation su e `eventName="GetDataAccess"` gli eventi S3 su `eventName="Get Object"` o. `eventName="Put Object"`

Di seguito è riportato un estratto di un CloudTrail evento cross-account di Lake Formation in cui è stato effettuato l'accesso ai dati in una posizione S3 registrata.

```
{
  "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"
   }
}
```

Il valore `lakeFormationRoleSessionName` chiave,`AWSLF-00-GL-111122223333-B8JSAjo5QA`, può essere unito al nome della sessione nella `principalId` chiave dell'evento S3. CloudTrail Di seguito è riportato un estratto dell'evento S3. CloudTrail Mostra la posizione del nome della sessione.

```
{
   "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"
     },
   ..............
   ..............
}
```

Il nome della sessione è formattato come segue:

```
AWSLF-<version-number>-<query-engine-code>-<account-id->-<suffix>
```

`version-number`

La versione di questo formato, attualmente`00`. Se il formato del nome della sessione cambia, sarà la versione successiva`01`.

`query-engine-code`

Indica l'entità che ha avuto accesso ai dati. I valori correnti sono:

`GL`	AWS GlueLavoro ETL
`AT`	Athena
`RE`	Amazon Redshift Spectrum

`account-id`

L'ID AWS dell'account che ha richiesto le credenziali a Lake Formation.

`suffix`

Una stringa generata casualmente.

Gestione delle autorizzazioni tra account utilizzando entrambi AWS Glue e Lake Formation

È possibile concedere l'accesso a più account alle risorse del Data Catalog e ai dati sottostanti utilizzando o. AWS Glue AWS Lake Formation

InAWS Glue, concedi autorizzazioni per più account creando o aggiornando una politica delle risorse di Data Catalog. In Lake Formation, concedi autorizzazioni per più account utilizzando il modello di autorizzazioni Lake Formation `GRANT/REVOKE` e il funzionamento dell'`Grant Permissions`API.

> **Tip**
>
> Ti consigliamo di affidarti esclusivamente alle autorizzazioni di Lake Formation per proteggere il tuo data lake.

Puoi visualizzare le sovvenzioni di Lake Formation su più account utilizzando la console Lake Formation o la console AWS Resource Access Manager (AWS RAM). Tuttavia, queste pagine della console non mostrano le autorizzazioni per più account concesse dalla politica delle risorse di AWS Glue Data Catalog. Allo stesso modo, puoi visualizzare le concessioni tra account nella politica delle risorse di Data Catalog utilizzando la pagina Impostazioni della AWS Glue console, ma quella pagina non mostra le autorizzazioni per più account concesse utilizzando Lake Formation.

Per assicurarti di non perdere nessuna sovvenzione durante la visualizzazione e la gestione delle autorizzazioni tra più account, Lake Formation AWS Glue richiede che tu esegua le seguenti azioni per indicare che sei a conoscenza e che stai autorizzando le sovvenzioni tra account sia da Lake Formation che. AWS Glue

Quando si concedono autorizzazioni su più account utilizzando la politica delle risorse di Data Catalog AWS Glue

Se il tuo account (account concedente o account produttore) non ha concesso concessioni tra account diversi che vengono utilizzate AWS RAM per condividere le risorse, puoi salvare una politica sulle risorse di Data Catalog come di consueto in. AWS Glue Tuttavia, se sono già state concesse sovvenzioni che prevedono la condivisione AWS RAM delle risorse, è necessario effettuare una delle seguenti operazioni per garantire che la politica di salvataggio delle risorse abbia esito positivo:

- Quando salvi la politica delle risorse nella pagina Impostazioni della AWS Glue console, la console emette un avviso che indica che le autorizzazioni nella politica si aggiungeranno a tutte le autorizzazioni concesse utilizzando la console Lake Formation. È necessario scegliere Procedi per salvare la politica.
- Quando si salva la politica delle risorse utilizzando l'operazione `glue:PutResourcePolicy` API, è necessario impostare il `EnableHybrid` campo su 'TRUE' (type = string). Il seguente esempio di codice mostra come eseguire questa operazione in Python.

```
import boto3
import json
```

```
REGION = 'us-east-2'
PRODUCER_ACCOUNT_ID = '123456789012'
CONSUMER_ACCOUNT_IDs = ['111122223333']

glue = glue_client = boto3.client('glue')

policy = {
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "Cataloguers",
            "Effect": "Allow",
            "Action": [
                "glue:*"
            ],
            "Principal": {
                "AWS": CONSUMER_ACCOUNT_IDs
            },
            "Resource": [
                f"arn:aws:glue:{REGION}:{PRODUCER_ACCOUNT_ID}:catalog",
                f"arn:aws:glue:{REGION}:{PRODUCER_ACCOUNT_ID}:database/*",
                f"arn:aws:glue:{REGION}:{PRODUCER_ACCOUNT_ID}:table/*/*"
            ]
        }
    ]
}

policy = json.dumps(policy)
glue.put_resource_policy(PolicyInJson=policy, EnableHybrid='TRUE')
```

Per ulteriori informazioni, consulta PutResourcePolicy Action (Python: put_resource_policy) nella Developer Guide.AWS Glue

Quando si concedono autorizzazioni per più account utilizzando il metodo delle risorse denominate Lake Formation

Se nel tuo account non è presente alcuna politica sulle risorse di Data Catalog, Lake Formation ti consente di procedere come al solito. Tuttavia, se esiste una politica in materia di risorse di Data Catalog, devi aggiungervi la seguente dichiarazione per consentire che le concessioni tra account abbiano esito positivo se effettuate con il metodo della risorsa denominata. <region>Sostituiscila con un nome di regione valido e <account-id>con l'ID del tuo AWS account.

```
{
  "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"
  ]
}
```

Senza questa dichiarazione aggiuntiva, la sovvenzione di Lake Formation ha esito positivo, ma viene bloccata e l'account del destinatario non può accedere alla risorsa concessa. AWS RAM

Important

Quando si utilizza il metodo di controllo degli accessi basato su tag Lake Formation (LF-TBAC) per concedere concessioni tra account, è necessario disporre di una politica delle risorse del Data Catalog con almeno le autorizzazioni specificate in. Prerequisiti

Consulta anche:

- Controllo dell'accesso ai metadati(per una discussione sul metodo Named Resource rispetto al metodo di controllo degli accessi basato su tag Lake Formation (LF-TBAC)).
- Visualizzazione di tabelle e database condivisi del catalogo dati
- Utilizzo delle impostazioni del catalogo dati sulla console nella Guida per gli sviluppatori AWS GlueAWS Glue
- Concessione dell'accesso a più account nella Guida per gli AWS Glue sviluppatori (ad esempio, le politiche relative alle risorse di Data Catalog)

Visualizzazione di tutte le sovvenzioni tra account utilizzando l'operazione API GetResourceShares

Se la tua azienda concede autorizzazioni per più account utilizzando sia una politica AWS Glue Data Catalog delle risorse che le sovvenzioni Lake Formation, l'unico modo per visualizzare tutte le sovvenzioni su più account in un unico posto è utilizzare l'operazione API. `glue:GetResourceShares`

Quando concedi le autorizzazioni di Lake Formation su più account utilizzando il metodo di risorsa denominato, AWS Resource Access Manager (AWS RAM) crea una politica delle risorse AWS Identity and Access Management (IAM) e la memorizza nel tuo AWS account. La politica concede le autorizzazioni necessarie per accedere alla risorsa. AWS RAM crea una politica delle risorse separata per ogni concessione tra account. È possibile visualizzare tutte queste politiche utilizzando l'operazione `glue:GetResourceShares` API.

> **Note**
>
> Questa operazione restituisce anche la politica delle risorse di Data Catalog. Tuttavia, se hai abilitato la crittografia dei metadati nelle impostazioni del Catalogo dati e non disponi dell'autorizzazione sulla AWS KMS chiave, l'operazione non restituirà la politica delle risorse del Catalogo dati.

Per visualizzare tutte le sovvenzioni concesse tra account

- Immettere il seguente comando AWS CLI .

```
aws glue get-resource-policies
```

Di seguito è riportato un esempio di politica delle risorse che AWS RAM crea e memorizza quando si concedono autorizzazioni su una tabella `t` nel database `db1` all' AWS account 1111-2222-3333.

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

```
          "glue:GetTable",
          "glue:GetTables",
          "glue:GetTableVersion",
          "glue:GetTableVersions",
          "glue:GetPartition",
          "glue:GetPartitions",
          "glue:BatchGetPartition",
          "glue:SearchTables"
        ],
       "Principal": {"AWS": [
         "111122223333"
       ]},
       "Resource": [
         "arn:aws:glue:<region>:111122223333:table/db1/t"
     ]
    }
  ]
}
```

Consulta anche:

- GetResourceShares Azione (Python: get_resource_policies) nella Guida per gli sviluppatori AWS Glue

Accesso e visualizzazione di tabelle e database condivisi del Data Catalog

Per l'amministratore del data lake e per i responsabili a cui sono state concesse le autorizzazioni, le risorse condivise con l'AWSaccount vengono visualizzate nel catalogo dati come se fossero risorse del tuo account. La console visualizza l'account proprietario della risorsa.

Puoi visualizzare le risorse condivise con il tuo account utilizzando la console Lake Formation. Puoi anche utilizzare la consoleAWS Resource Access Manager (AWS RAM) per visualizzare sia le risorse condivise con il tuo account sia le risorse che hai condiviso con altriAWS account utilizzando il metodo della risorsa denominata.

Important

Quando qualcuno utilizza il metodo della risorsa denominata per concedere autorizzazioni tra account su una risorsa del catalogo dati al tuo account o alla tuaAWS organizzazione, Lake Formation utilizza il servizioAWS Resource Access Manager (AWS RAM) per condividere la risorsa. Se il tuo account appartiene alla stessaAWS organizzazione dell'account concedente, la risorsa condivisa è immediatamente disponibile.
Tuttavia, se l'account non appartiene alla stessa organizzazione,AWS RAM invia un invito all'account per accettare o rifiutare la condivisione di risorse. Quindi, per rendere disponibile la risorsa condivisa, l'amministratore del data lake dell'account deve utilizzare laAWS RAM console o la CLI per accettare l'invito.
La console di Lake Formation visualizza un avviso se c'è un invito alla condivisione diAWS RAM risorse in attesa di essere accettato. Solo gli utenti autorizzati a visualizzareAWS RAM gli inviti ricevono l'avviso.

Consulta anche:

- Condivisione di tabelle e database del Data Catalog tra account AWS
- Condivisione dei dati tra account in Lake Formation
- Accesso ai dati sottostanti di una tabella condivisa
- Controllo dell'accesso ai metadati(per informazioni sul metodo della risorsa denominato rispetto al metodo LF-TBAC per la condivisione delle risorse.)

Argomenti

Accettazione di un invito alla condivisione delle risorse daAWS RAM

Se una risorsa del catalogo dati è condivisa con il tuoAWS account e il tuo account non è nella stessaAWS organizzazione dell'account di condivisione, non puoi accedere alla risorsa condivisa finché non accetti un invito alla condivisione di risorse daAWS Resource Access Manager (AWS

RAM). In qualità di amministratore del data lake, devi primaAWS RAM richiedere gli inviti in sospeso e poi accettarli.

Puoi usare laAWS RAM console, l'API oAWS Command Line Interface (AWS CLI) per visualizzare e accettare gli inviti.

Per visualizzare e accettare un invito alla condivisione di risorse daAWS RAM (console)

1. Assicurati di disporre delle autorizzazioniAWS Identity and Access Management (IAM) necessarie per visualizzare e accettare le autorizzazioni alla condivisione delle risorse.

   Per informazioni sulle politiche IAM suggerite per gli amministratori di data lake, consultathe section called “Autorizzazioni di amministratore di Data Lake”.
2. Segui le istruzioni contenute in Accettazione e rifiuto degli inviti nella Guida per l'AWS RAMutente.

Per visualizzare e accettare un invito alla condivisione di risorse daAWS RAM (AWSCLI)

1. Assicurati di disporre delle autorizzazioniAWS Identity and Access Management (IAM) necessarie per visualizzare e accettare le autorizzazioni alla condivisione delle risorse.

   Per informazioni sulle politiche IAM suggerite per gli amministratori di data lake, consultathe section called “Autorizzazioni di amministratore di Data Lake”.
2. Inserisci il comando seguente per visualizzare le richieste di condivisione delle risorse in sospeso.

   ```
   aws ram get-resource-share-invitations
   ```

   L'output visualizzato dovrebbe essere simile al seguente.

   ```
   {
       "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"
        }
    ]
}
```

Nota lo stato di`PENDING`.

3. Copia il valore della`resourceShareInvitationArn` chiave negli appunti.
4. Incolla il valore nel seguente comando<invitation-arn>, sostituiscilo e inserisci il comando.

```
aws ram accept-resource-share-invitation --resource-share-invitation-
arn <invitation-arn>
```

L'output visualizzato dovrebbe essere simile al seguente.

```
{
    "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"
        }
    ]
}
```

Nota lo stato di`ACCEPTED`.

Visualizzazione di tabelle e database condivisi del catalogo dati

Puoi visualizzare le risorse condivise con il tuo account utilizzando la console di Lake Formation o laAWS CLI. Puoi anche utilizzare la consoleAWS Resource Access Manager (AWS RAM) o la

CLI per visualizzare sia le risorse condivise con il tuo account sia le risorse che hai condiviso con altriAWS account.

Per visualizzare le risorse condivise utilizzando la console Lake Formation

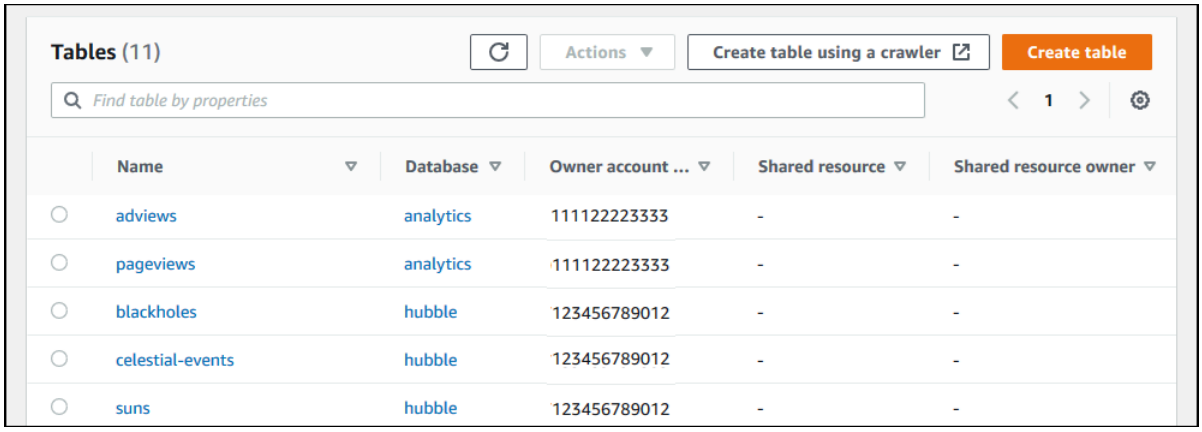
1. Aprire la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.

   Accedi come amministratore del data lake o utente a cui sono state concesse le autorizzazioni su una tabella condivisa.

2. Per visualizzare le risorse condivise con il tuoAWS account, completare una delle operazioni seguenti:

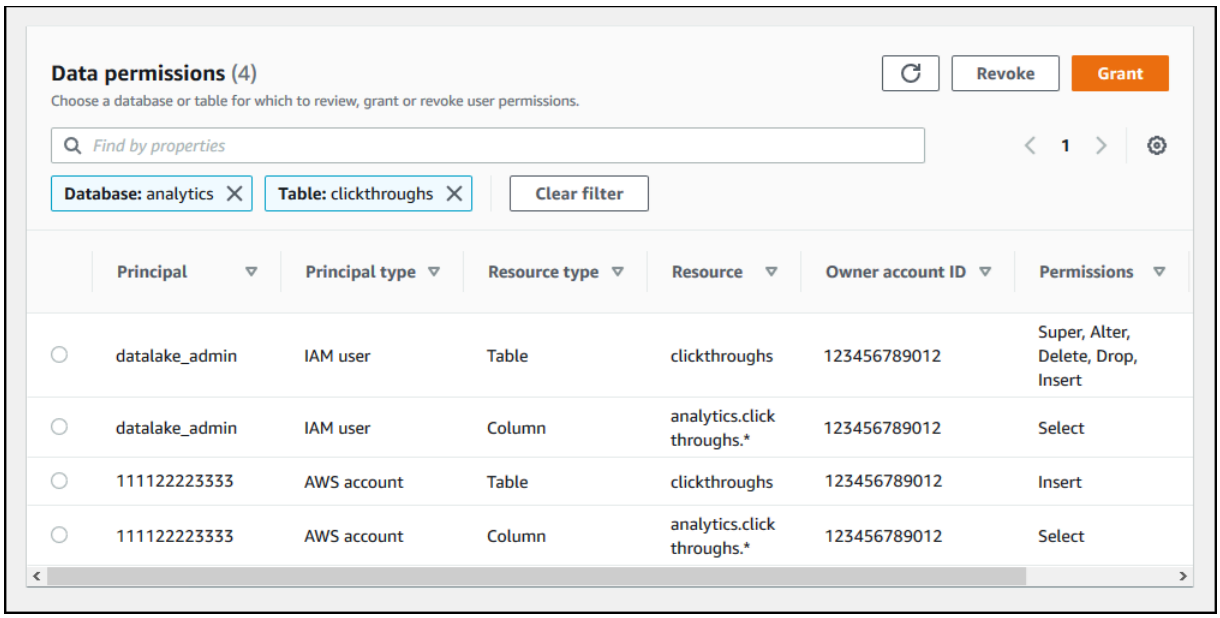
   - Per visualizzare le tabelle condivise con il tuo account, nel riquadro di navigazione, scegli Tabelle.
   - Per visualizzare i database condivisi con il tuo account, nel riquadro di navigazione, scegli Database.

   La console mostra un elenco di database o tabelle sia nel tuo account che condivisi con il tuo account. Per le risorse condivise con il tuo account, la console visualizza l'ID dell'AWSaccount del proprietario nella colonna ID account proprietario (la terza colonna nella schermata seguente).



3. Per visualizzare le risorse che hai condiviso con altriAWS account o organizzazioni, nel riquadro di navigazione, scegli Autorizzazioni dati.

   Le risorse che hai condiviso sono elencate nella pagina Autorizzazioni dati con il numero di conto esterno mostrato nella colonna Principale, come mostrato nell'immagine seguente.



Per visualizzare le risorse condivise tramite laAWS RAM console

1. Assicurati di disporre delle autorizzazioniAWS Identity and Access Management (IAM) necessarie per visualizzare le risorse condiviseAWS RAM.

   Come minimo, devi disporre dell'autorizzazione`ram:ListResources`. Questa autorizzazione è inclusa nella policy gestita di AWS`AWSLakeFormationCrossAccountManager`.
2. Accedere aAWS Management Console e aprire laAWS RAM console all'indirizzo https://console.aws.amazon.com/ram.
3. Completa una delle seguenti operazioni:

   - Per visualizzare le risorse che hai condiviso, nel riquadro di navigazione, in Condivise da me, scegli Risorse condivise.
   - Per visualizzare le risorse condivise con te, nel riquadro di navigazione, in Condivise con me, scegli Risorse condivise.

Creazione di collegamenti alle risorse

I link alle risorse sono oggetti del Data Catalog che sono collegamenti a database e tabelle di metadati, in genere a database e tabelle condivisi di altri account. AWS Aiutano a consentire l'accesso tra account diversi ai dati nel data lake in tutte le regioni. AWS

> **Note**
>
> Lake Formation supporta l'interrogazione delle tabelle del Data Catalog tra AWS le regioni. È possibile accedere ai database e alle tabelle del Data Catalog da qualsiasi AWS regione creando collegamenti di risorse in quelle aree che puntano a database e tabelle condivisi in diverse regioni.

Argomenti

Come funzionano i link alle risorse in Lake Formation

Un collegamento a una risorsa è un oggetto del catalogo dati che è un collegamento a un database o a una tabella locale o condivisa. Dopo aver creato un link di risorsa a un database o a una tabella, è possibile utilizzare il nome del link di risorsa ovunque si utilizzi il nome del database o della tabella. Oltre alle tabelle di tua proprietà o alle tabelle condivise con te, i link alle risorse delle tabelle vengono restituiti **`glue:GetTables()`** e vengono visualizzati come voci nella pagina Tabelle della console Lake Formation. I collegamenti alle risorse ai database agiscono in modo simile.

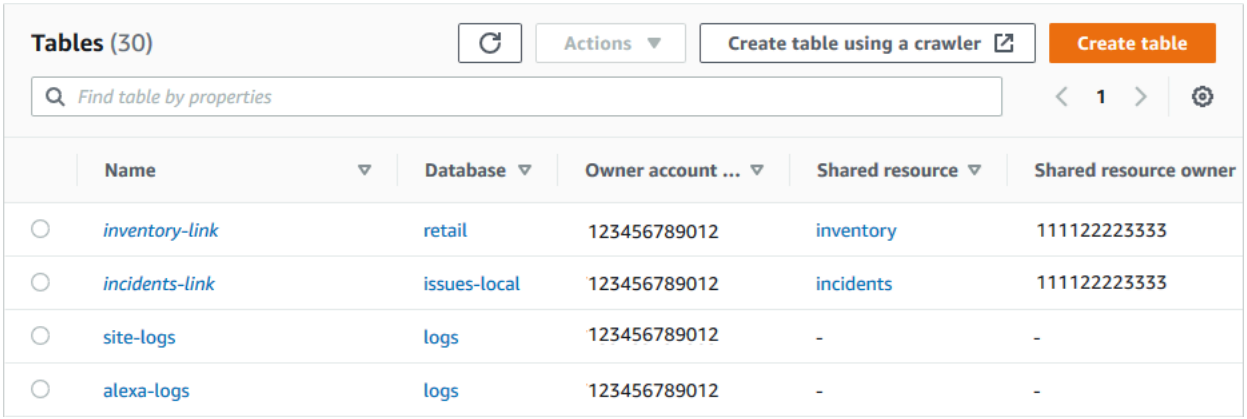
La creazione di un collegamento di risorsa a un database o a una tabella consente di effettuare le seguenti operazioni:

- Assegna un nome diverso a un database o a una tabella nel tuo Data Catalog. Ciò è particolarmente utile se AWS account diversi condividono database o tabelle con lo stesso nome o se più database dell'account hanno tabelle con lo stesso nome.
- Accedi ai database e alle tabelle del Data Catalog da qualsiasi AWS regione creando collegamenti alle risorse in quelle aree che puntano al database e alle tabelle in un'altra regione. Puoi eseguire query in qualsiasi regione con questi link alle risorse utilizzando Athena, Amazon EMR ed AWS Glue eseguire job ETL Spark, senza copiare i dati di origine né i metadati in Glue Data Catalog.
- Utilizza AWS servizi integrati come Amazon Athena Amazon Redshift Spectrum per eseguire query che accedono a database o tabelle condivisi. Alcuni servizi integrati non possono accedere

direttamente a database o tabelle tra account. Tuttavia, possono accedere ai link delle risorse presenti nel tuo account ai database e alle tabelle di altri account.

> **Note**
>
> Non è necessario creare un collegamento a una risorsa per fare riferimento a un database o a una tabella condivisa negli script di AWS Glue estrazione, trasformazione e caricamento (ETL). Tuttavia, per evitare ambiguità quando più AWS account condividono un database o una tabella con lo stesso nome, è possibile creare e utilizzare un collegamento alla risorsa o specificare l'ID del catalogo quando si richiamano le operazioni ETL.

L'esempio seguente mostra la pagina Tabelle della console di Lake Formation, che elenca due collegamenti a risorse. I nomi dei link alle risorse sono sempre visualizzati in corsivo. Ogni collegamento alla risorsa viene visualizzato insieme al nome e al proprietario della risorsa condivisa collegata. In questo esempio, un amministratore di Data Lake nell'AWSaccount 1111-2222-3333 ha condiviso le `incidents` tabelle `inventory` and con l'account 1234-5678-9012. Un utente di quell'account ha quindi creato collegamenti alle risorse a tali tabelle condivise.



Di seguito sono riportate le note e le restrizioni relative ai collegamenti alle risorse:

- I link alle risorse sono necessari per consentire a servizi integrati come Athena e Redshift Spectrum di interrogare i dati sottostanti delle tabelle condivise. Le query in questi servizi integrati sono costruite sulla base dei nomi dei link alle risorse.
- Supponendo che l'impostazione Usa solo il controllo di accesso IAM per le nuove tabelle in questo database sia disattivata per il database che lo contiene, solo l'utente principale che ha creato un collegamento alla risorsa può visualizzarlo e accedervi. Per consentire agli altri responsabili del tuo account di accedere a un collegamento a una risorsa, concedi

l'DESCRIBEautorizzazione al collegamento. Per consentire ad altri di eliminare un link a una risorsa, concedi l'DROPautorizzazione. Gli amministratori di Data Lake possono accedere a tutti i link alle risorse presenti nell'account. Per eliminare un collegamento a una risorsa creato da un altro principale, l'amministratore del data lake deve prima concedersi l'DROPautorizzazione sul collegamento alla risorsa. Per ulteriori informazioni, consulta Riferimento alle autorizzazioni di Lake Formation.

> **Important**
>
> La concessione delle autorizzazioni per un collegamento a una risorsa non concede le autorizzazioni per il database o la tabella di destinazione (collegati). È necessario concedere le autorizzazioni sulla destinazione separatamente.

- Per creare un collegamento a una risorsa, è necessaria l'CREATE_DATABASEautorizzazione CREATE_TABLE o la Lake Formation, nonché l'autorizzazione `glue:CreateTable` or `glue:CreateDatabase` AWS Identity and Access Management (IAM).
- Puoi creare collegamenti a risorse locali (di proprietà) del Data Catalog, nonché a risorse condivise con il tuo AWS account.
- Quando crei un link a una risorsa, non viene eseguito alcun controllo per verificare se la risorsa condivisa di destinazione esiste o se disponi di autorizzazioni per più account sulla risorsa. Ciò consente di creare il collegamento alla risorsa e la risorsa condivisa in qualsiasi ordine.
- Se si elimina un collegamento a una risorsa, la risorsa condivisa collegata non viene eliminata. Se si elimina una risorsa condivisa, i link delle risorse a tale risorsa non vengono eliminati.
- È possibile creare catene di collegamenti di risorse. Tuttavia, non è utile farlo, perché le API seguono solo il primo collegamento alla risorsa.

> **Consulta anche:**
>
> - Concessione e revoca delle autorizzazioni per le risorse del Data Catalog

Creazione di un collegamento di risorsa a una tabella condivisa del catalogo dati

Puoi creare un link di risorsa a una tabella condivisa in qualsiasi AWS regione utilizzando la AWS Lake Formation console, l'API o AWS Command Line Interface (). AWS CLI

Per creare un collegamento di risorsa a una tabella condivisa (console)

1. Apri la AWS Lake Formation console all'indirizzo https://console.aws.amazon.com/lakeformation/. Accedi come preside che dispone dell'`CREATE_TABLE`autorizzazione Lake Formation sul database per contenere il link alla risorsa.
2. Nel pannello di navigazione, scegli Tabelle, quindi seleziona Crea tabella.
3. Nella pagina Crea tabella, scegli il riquadro Resource Link, quindi fornisci le seguenti informazioni:

   Nome del link alla risorsa

   Immettete un nome che rispetti le stesse regole del nome della tabella. Il nome può essere lo stesso della tabella condivisa di destinazione.

   Database

   Il database nel Data Catalog locale che deve contenere il collegamento alla risorsa.

   Regione proprietaria della tabella condivisa

   Se stai creando il link alla risorsa in un'altra regione, seleziona la regione della tabella condivisa di destinazione.

   Tabella condivisa

   Seleziona una tabella condivisa dall'elenco o inserisci un nome di tabella locale (di proprietà) o condivisa.

   L'elenco contiene tutte le tabelle condivise con il tuo account. Annota il database e l'ID dell'account del proprietario elencati in ogni tabella. Se non vedi una tabella che sai è stata condivisa con il tuo account, controlla quanto segue:

   - Se non sei un amministratore del data lake, verifica che l'amministratore del data lake ti abbia concesso le autorizzazioni Lake Formation sulla tabella.
   - Se sei un amministratore del data lake e il tuo account non fa parte della stessa AWS organizzazione dell'account concedente, assicurati di aver accettato l'invito alla

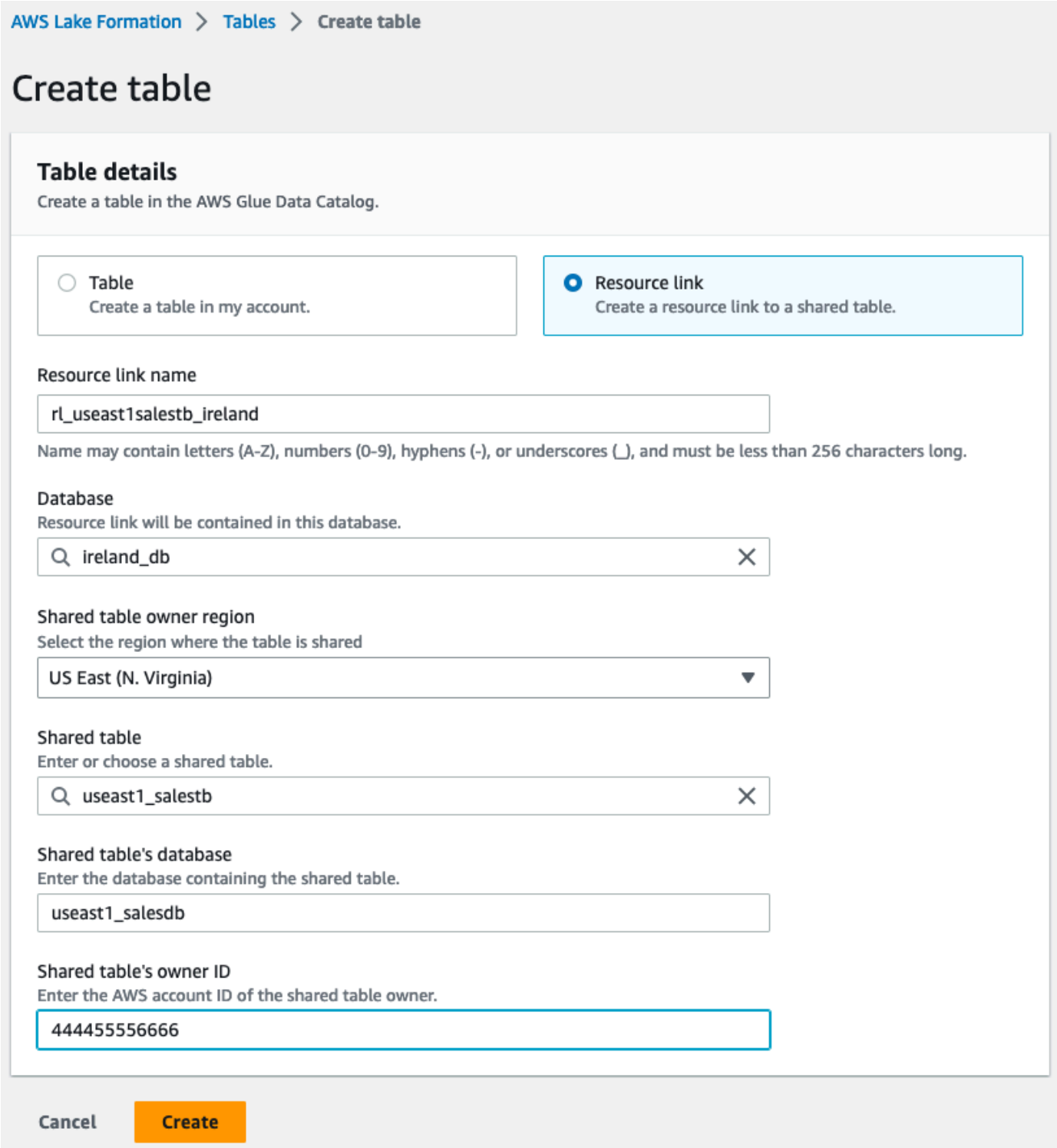
condivisione delle risorse AWS Resource Access Manager (AWS RAM) per la tabella. Per ulteriori informazioni, consulta Accettazione di un invito alla condivisione delle risorse daAWS RAM.

Database della tabella condivisa

Se hai selezionato una tabella condivisa dall'elenco, questo campo viene popolato con il database della tabella condivisa nell'account esterno. Altrimenti, inserisci un database locale (per un collegamento di risorse a una tabella locale) o il database della tabella condivisa nell'account esterno.

Proprietario della tabella condivisa

Se hai selezionato una tabella condivisa dall'elenco, questo campo viene compilato con l'ID dell'account proprietario della tabella condivisa. Altrimenti, inserisci l'ID AWS del tuo account (per il collegamento di una risorsa a una tabella locale) o l'ID dell'AWSaccount che ha condiviso la tabella.



4. Scegli Crea per creare il link alla risorsa.

   È quindi possibile visualizzare il nome del collegamento alla risorsa nella colonna Nome della pagina Tabelle.

5. (Facoltativo) Concedi l'`DESCRIBE`autorizzazione Lake Formation sul link della risorsa ai responsabili che devono essere in grado di visualizzare il collegamento e accedere alla tabella di destinazione.

Per creare un collegamento di risorsa a una tabella condivisa nella stessa regione () AWS CLI

1. Utilizzare un comando simile al seguente:

```
aws glue create-table --database-name myissues --table-input
 '{"Name":"my_customers","TargetTable":
{"CatalogId":"111122223333","DatabaseName":"issues","Name":"customers"}}'
```

   Questo comando crea un link di risorsa denominato `my_customers` alla tabella condivisa`customers`, che si trova `issues` nel database dell'AWSaccount 1111-2222-3333. Il collegamento alla risorsa è memorizzato nel database locale. `myissues`

2. (Facoltativo) Concedi l'`DESCRIBE`autorizzazione Lake Formation sul link della risorsa ai responsabili che devono essere in grado di visualizzare il collegamento e accedere alla tabella di destinazione.

Per creare un collegamento di risorsa a una tabella condivisa in un'altra regione () AWS CLI

1. Utilizzare un comando simile al seguente:

```
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"
        }
    }
}‘
```

   Questo comando crea un link di risorsa denominato `rl_useast1salestb_irelan`d nella regione Europa (Irlanda) alla tabella condivisa`useast1_salestb`, che si trova nel database

dell'AWSaccount 444455556666 `useast1_salesdb` nella regione Stati Uniti orientali (Virginia settentrionale). Il collegamento alla risorsa è memorizzato nel database locale. `ireland_db`

2. Concedi l'`DESCRIBE`autorizzazione a Lake Formation ai responsabili che devono essere in grado di visualizzare il link e accedere alla destinazione del link tramite il link.

> Consulta anche:
>
> - Come funzionano i link alle risorse in Lake Formation
> - `DESCRIBE`

Creazione di un collegamento di risorsa a un database Data Catalog condiviso

È possibile creare un collegamento di risorsa a un database condiviso utilizzando la AWS Lake Formation console, l'API o AWS Command Line Interface ()AWS CLI.

Per creare un collegamento di risorsa a un database condiviso (console)

1. Apri la AWS Lake Formation console all'indirizzo https://console.aws.amazon.com/lakeformation/. Accedi come amministratore del data lake o come creatore di database.

   Un creatore di database è un responsabile a cui è stata concessa l'`CREATE_DATABASE`autorizzazione Lake Formation.

2. Nel riquadro di navigazione, scegli Database, quindi scegli Crea database.
3. Nella pagina Crea database, scegli il riquadro Resource Link, quindi fornisci le seguenti informazioni:

   Nome del link alla risorsa

   Immettete un nome che rispetti le stesse regole del nome del database. Il nome può essere lo stesso del database condiviso di destinazione.

   Regione del proprietario del database condiviso

   Se stai creando il link alla risorsa in un'altra regione, seleziona la regione del database condiviso di destinazione.

Database condiviso

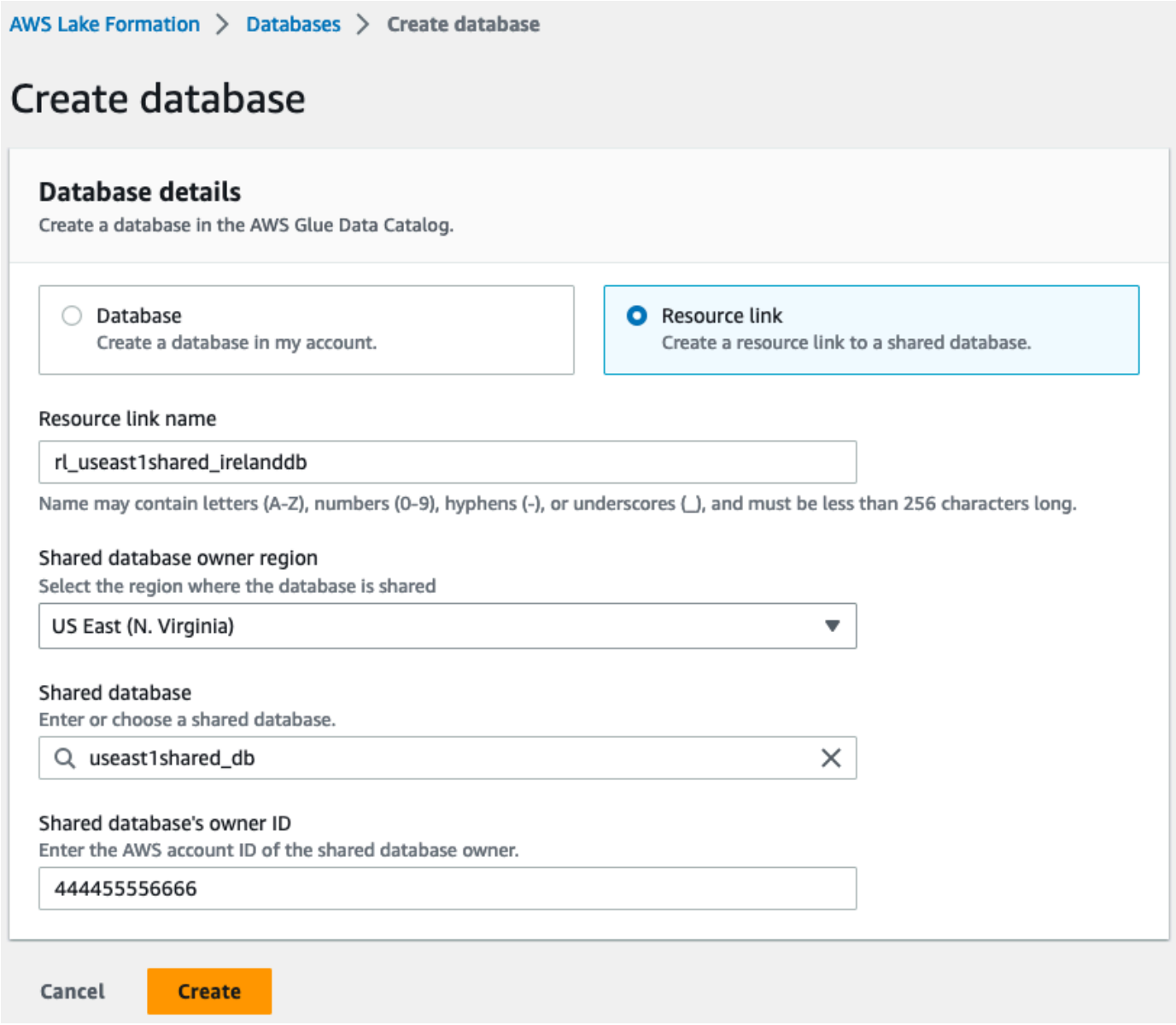
Scegli un database dall'elenco o inserisci un nome di database locale (di proprietà) o condiviso.

L'elenco contiene tutti i database condivisi con il tuo account. Annota l'ID dell'account del proprietario elencato in ogni database. Se non vedi un database che sai è stato condiviso con il tuo account, controlla quanto segue:

- Se non sei un amministratore del data lake, verifica che l'amministratore del data lake ti abbia concesso le autorizzazioni Lake Formation sul database.
- Se sei un amministratore del data lake e il tuo account non fa parte della stessa AWS organizzazione dell'account concedente, assicurati di aver accettato l'invito AWS Resource Access Manager (AWS RAM) alla condivisione delle risorse per il database. Per ulteriori informazioni, consulta Accettazione di un invito alla condivisione delle risorse daAWS RAM.

Proprietario del database condiviso

Se hai selezionato un database condiviso dall'elenco, questo campo viene compilato con l'ID dell'account proprietario del database condiviso. Altrimenti, inserisci l'ID AWS del tuo account (per un collegamento di risorsa a un database locale) o l'ID dell'AWSaccount che ha condiviso il database.



4. Scegli Crea per creare il link alla risorsa.

   È quindi possibile visualizzare il nome del collegamento alla risorsa nella colonna Nome della pagina Database.

5. (Facoltativo) Concedi a Lake Formation l'`DESCRIBE`autorizzazione per il collegamento alla risorsa ai responsabili della regione Europa (Irlanda) che devono essere in grado di visualizzare il collegamento e accedere al database di destinazione.

Per creare un collegamento di risorsa a un database condiviso nella stessa regione () AWS CLI

1. Utilizzare un comando simile al seguente:

```
aws glue create-database --database-input '{"Name":"myissues","TargetDatabase":
{"CatalogId":"111122223333","DatabaseName":"issues"}}'
```

   Questo comando crea un collegamento a una risorsa denominato `myissues` al database condivisoissues, che si trova nell'AWSaccount 1111-2222-3333.

2. (Facoltativo) Concedi l'`DESCRIBE`autorizzazione a Lake Formation ai responsabili del link alla risorsa che devono essere in grado di visualizzare il collegamento e accedere al database di destinazione.

Per creare un collegamento di risorsa a un database condiviso in un'altra regione () AWS CLI

1. Utilizzare un comando simile al seguente:

```
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"
       }
    }
}'
```

   Questo comando crea un collegamento di risorsa denominato `rl_useast1shared_irelanddb` nell'AWSaccount 111122223333 nella regione Europa (Irlanda) al database condivisouseast1shared_db, che si trova nell'AWSaccount 444455556666 nella regione Stati Uniti orientali (Virginia settentrionale).

2. Concedi il `DESCRIBE` permesso di Lake Formation ai dirigenti della regione Europa (Irlanda) che devono essere in grado di visualizzare il link e accedere alla destinazione del link tramite il link.

Consulta anche:

- Come funzionano i link alle risorse in Lake Formation
- `DESCRIBE`

Gestione dei link alle risorse nelle API AWS Glue

Le tabelle seguenti spiegano come le API AWS Glue Data Catalog gestiscono i collegamenti alle risorse di database e tabelle. Per tutte le operazioni `Get*` API, vengono restituiti solo i database e le tabelle per i quali il chiamante dispone delle autorizzazioni. Inoltre, quando si accede a un database o a una tabella di destinazione tramite un collegamento a una risorsa, è necessario disporre delle autorizzazioni sia AWS Identity and Access Management (IAM) che di Lake Formation sia sulla destinazione che sul collegamento alla risorsa. L'autorizzazione di Lake Formation richiesta per i collegamenti alle risorse è`DESCRIBE`. Per ulteriori informazioni, consulta `DESCRIBE`.

Operazioni dell'API del database

API operation (Operazione API)	Gestione dei collegamenti alle risorse
`CreateDatabase`	Se il database è un collegamento di risorse, crea il collegamento alla risorsa al database di destinazione designato.
`UpdateDatabase`	Se il database designato è un collegamento a una risorsa, segue il collegamento e aggiorna il database di destinazione. Se è necessario modificare il collegamento alla risorsa per collegarsi a un database diverso, è necessario eliminarlo e crearne uno nuovo.
`DeleteDatabase`	Elimina il link alla risorsa. Non elimina il database collegato (di destinazione).
`GetDatabase`	Se il chiamante dispone delle autorizzazioni sulla destinazione, segue il link per restituire le proprietà della destinazione. Altrimenti, restituisce le proprietà del link.
`GetDatabases`	Restituisce un elenco di database, inclusi i collegamenti alle risorse. Per ogni collegamento di risorsa nel set di risultati, l'operazione

API operation (Operazione API)	Gestione dei collegamenti alle risorse
	segue il collegamento per ottenere le proprietà della destinazione del collegamento. Devi specificare `ResourceShareType = ALL` per vedere i database condivisi con il tuo account.

Tabella delle operazioni API

API operation (Operazione API)	Gestione dei link alle risorse
`CreateTable`	Se il database è un collegamento a una risorsa, segue il collegamento al database e crea una tabella nel database di destinazione. Se la tabella è un collegamento di risorse, l'operazione crea il collegamento alle risorse nel database designato. La creazione di un collegamento alle risorse della tabella tramite un collegamento alle risorse del database non è supportata.
`UpdateTable`	Se la tabella o il database designato è un collegamento a una risorsa, aggiorna la tabella di destinazione. Se sia la tabella che il database sono collegamenti a risorse, l'operazione ha esito negativo.
`DeleteTable`	Se il database designato è un collegamento a una risorsa, segue il collegamento ed elimina la tabella o il collegamento alle risorse della tabella nel database di destinazione. Se la tabella è un collegamento a una risorsa, l'operazione elimina il collegamento alle risorse della tabella nel database designato. L'eliminazione di un collegamento alle risorse della tabella non elimina la tabella di destinazione.
`BatchDeleteTable`	Come `DeleteTable` .
`GetTable`	Se il database designato è un collegamento di risorse, segue il collegamento al database e restituisce la tabella o il collegamento alle risorse della tabella dal database di destinazione. Altrimenti,

API operation (Operazione API)	Gestione dei link alle risorse
	se la tabella è un collegamento a una risorsa, l'operazione segue il collegamento e restituisce le proprietà della tabella di destinazione.
`GetTables`	Se il database designato è un collegamento a una risorsa, segue il collegamento al database e restituisce le tabelle e i collegamenti alle risorse della tabella dal database di destinazione. Se il database di destinazione è un database condiviso di un altro AWS account, l'operazione restituisce solo le tabelle condivise in quel database. Non segue i collegamenti alle risorse della tabella nel database di destinazione. Altrimenti, se il database designato è un database locale (di proprietà), l'operazione restituisce tutte le tabelle del database locale e segue ogni collegamento alle risorse della tabella per restituire le proprietà della tabella di destinazione.
`SearchTables`	Restituisce tabelle e collegamenti alle risorse delle tabelle. Non segue i link per restituire le proprietà della tabella di destinazione. Devi specificare `ResourceShareType = ALL` per vedere le tabelle condivise con il tuo account.
`GetTableVersion`	Come `GetTable`.
`GetTableVersions`	Come `GetTable`.
`DeleteTableVersion`	Come `DeleteTable` .
`BatchDeleteTableVersion`	Come `DeleteTable` .

Operazioni dell'API di partizione

API operation (Operazione API)	Gestione dei link alle risorse
`CreatePartition`	Se il database designato è un collegamento di risorse, segue il collegamento al database e crea una partizione nella tabella designata nel database di destinazione. Se la tabella è un

API operation (Operazione API)	Gestione dei link alle risorse
	collegamento di risorse, l'operazione segue il collegamento alle risorse e crea la partizione nella tabella di destinazione. La creazione di una partizione tramite un collegamento alle risorse della tabella e un collegamento alle risorse del database non è supportata.
`BatchCreatePartition`	Come `CreatePartition` .
`UpdatePartition`	Se il database designato è un collegamento a una risorsa, segue il collegamento al database e aggiorna la partizione nella tabella designata nel database di destinazione. Se la tabella è un collegamento a una risorsa, l'operazione segue il collegamento alla risorsa e aggiorna la partizione nella tabella di destinazione. L'aggiornamento di una partizione tramite un collegamento alle risorse della tabella e un collegamento alle risorse del database non è supportato.
`DeletePartition`	Se il database designato è un collegamento di risorse, segue il collegamento al database ed elimina la partizione nella tabella designata nel database di destinazione. Se la tabella è un collegamento di risorse, l'operazione segue il collegamento alla risorsa ed elimina la partizione nella tabella di destinazione. L'eliminazione di una partizione tramite un collegamento alle risorse della tabella e un collegamento alle risorse del database non è supportata.
`BatchDeletePartition`	Come `DeletePartition` .

API operation (Operazione API)	Gestione dei link alle risorse
`GetPartition`	Se il database designato è un collegamento a una risorsa, segue il collegamento al database e restituisce le informazioni sulla partizion e dalla tabella designata. Altrimenti, se la tabella è un collegame nto a una risorsa, l'operazione segue il collegamento e restituisce informazioni sulla partizione. Se sia la tabella che il database sono collegamenti a risorse, restituisce un set di risultati vuoto.
`GetPartitions`	Se il database designato è un collegamento di risorse, segue il collegamento al database e restituisce le informazioni sulla partizione per tutte le partizioni nella tabella designata. Altriment i, se la tabella è un collegamento di risorse, l'operazione segue il collegamento e restituisce informazioni sulla partizione. Se sia la tabella che il database sono collegamenti a risorse, restituisce un set di risultati vuoto.
`BatchGetPartition`	Come `GetPartition` .

Funzioni definite dall'utente, operazioni API

API operation (Operazione API)	Gestione dei link alle risorse
(Tutte le operazioni API)	Se il database è un collegamento a una risorsa, segue il collegame nto alla risorsa ed esegue l'operazione sul database di destinazi one.

Consulta anche:

- Come funzionano i link alle risorse in Lake Formation

Accesso alle tabelle in tutte le regioni

Lake Formation supporta l'interrogazione delle tabelle del Data Catalog tra AWS le regioni. Puoi accedere ai dati in una regione da altre regioni utilizzando Amazon Athena, Amazon EMR ed AWS Glue ETL creando collegamenti a risorse in altre regioni che puntano ai database e alle tabelle di origine. Con l'accesso alle tabelle tra regioni, puoi accedere ai dati di tutte le regioni senza copiare i dati o i metadati sottostanti nel catalogo dati.

Ad esempio, puoi condividere un database o una tabella in un account produttore con un account consumatore nella Regione A. Dopo aver accettato l'invito alla condivisione delle risorse nella Regione A, l'amministratore del data lake dell'account consumer può creare collegamenti alle risorse condivise nella Regione B. L'amministratore dell'account consumer può concedere le autorizzazioni sulla risorsa condivisa ai responsabili IAM di quell'account nella Regione A e può concedere le autorizzazioni per il collegamento alle risorse nella Regione B. Utilizzando il link alla risorsa, i principali nell'account consumer può interrogare i dati condivisi dalla regione B.

Puoi anche ospitare l'origine dati Amazon S3 nella Regione A in un account produttore e registrare la posizione dei dati in un account centrale nella Regione B. Puoi creare risorse Data Catalog nell'account centrale, impostare le autorizzazioni di Lake Formation e condividere dati con i consumatori nel tuo account o con account esterni nella Regione B. La funzionalità interregionale consente agli utenti di accedere a queste tabelle del Catalogo dati dalla Regione C utilizzando collegamenti alle risorse.

Utilizzando questa funzionalità, è possibile interrogare i database federati in Apache Hive Metastores tra le regioni e anche unire le tabelle nella regione locale con le tabelle di un'altra regione durante l'esecuzione delle query.

Lake Formation supporta le seguenti funzionalità con l'accesso alle tabelle tra regioni:

- Controllo degli accessi basato su tag LF
- Autorizzazioni di controllo degli accessi granulari
- Operazioni di scrittura sul database o sulla tabella condivisa con le autorizzazioni appropriate
- Condivisione dei dati tra account a livello di account e direttamente con i principali IAM

Gli utenti non amministrativi con `Create_Database` e `Create_Table` autorizzazioni possono creare collegamenti di risorse tra regioni.

> **Note**
>
> Puoi creare collegamenti a risorse interregionali in qualsiasi regione e accedere ai dati senza applicare le autorizzazioni di Lake Formation. Per i dati di origine in Amazon S3 che non sono registrati con Lake Formation, l'accesso è determinato dalle policy di autorizzazione IAM per Amazon S3 e dalle operazioni AWS Glue.

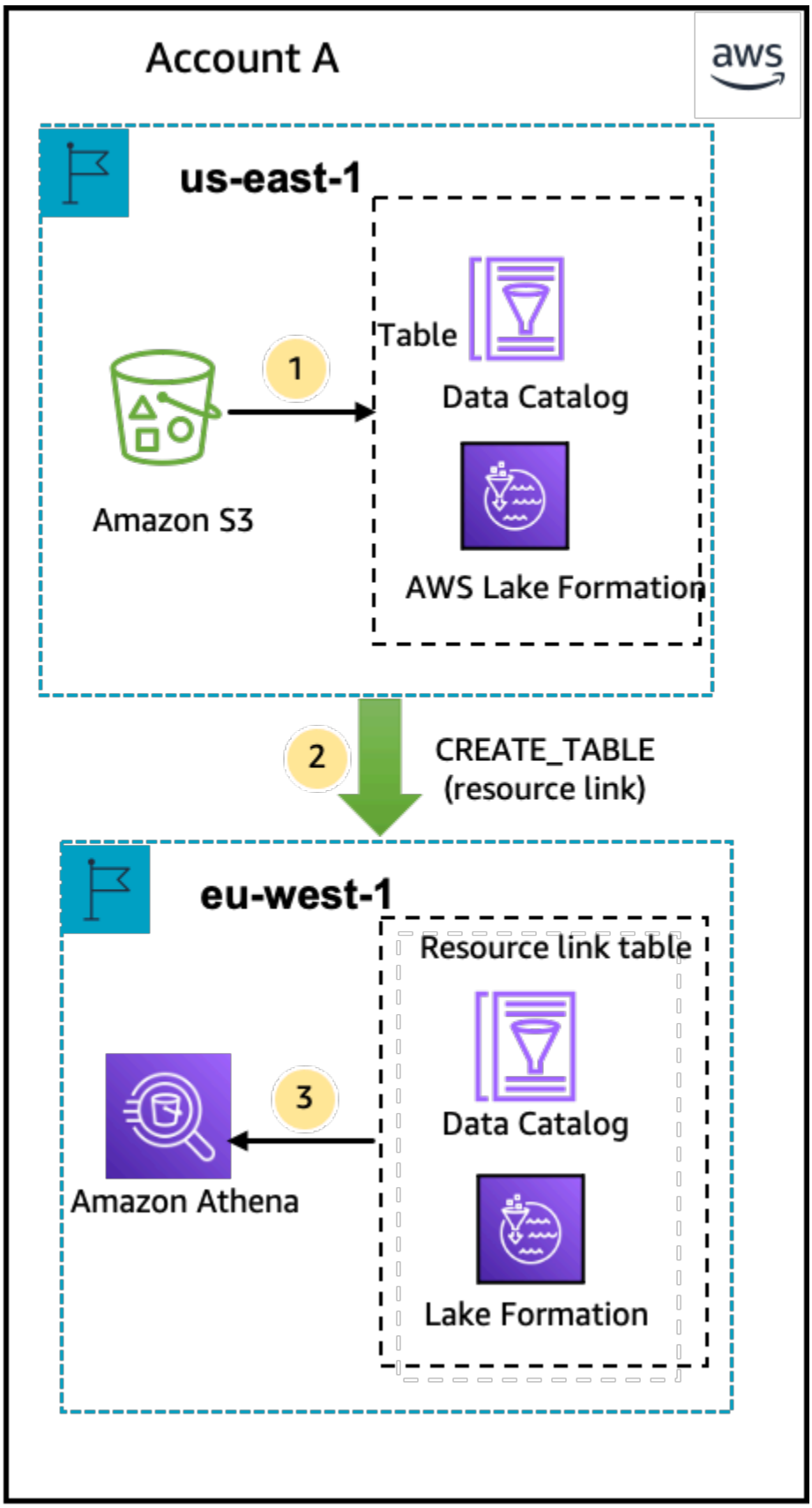
Per le limitazioni, consulta Limitazioni di accesso ai dati tra regioni.

Flussi di lavoro

I seguenti diagrammi mostrano i flussi di lavoro per l'accesso ai dati tra AWS le regioni dallo stesso AWS account e da un account esterno.

Flusso di lavoro per accedere alle tabelle condivise all'interno dello stesso account AWS

Nel diagramma seguente, i dati vengono condivisi con un utente dello stesso AWS account nella regione Stati Uniti orientali (Virginia settentrionale) e l'utente richiede i dati condivisi dalla regione 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

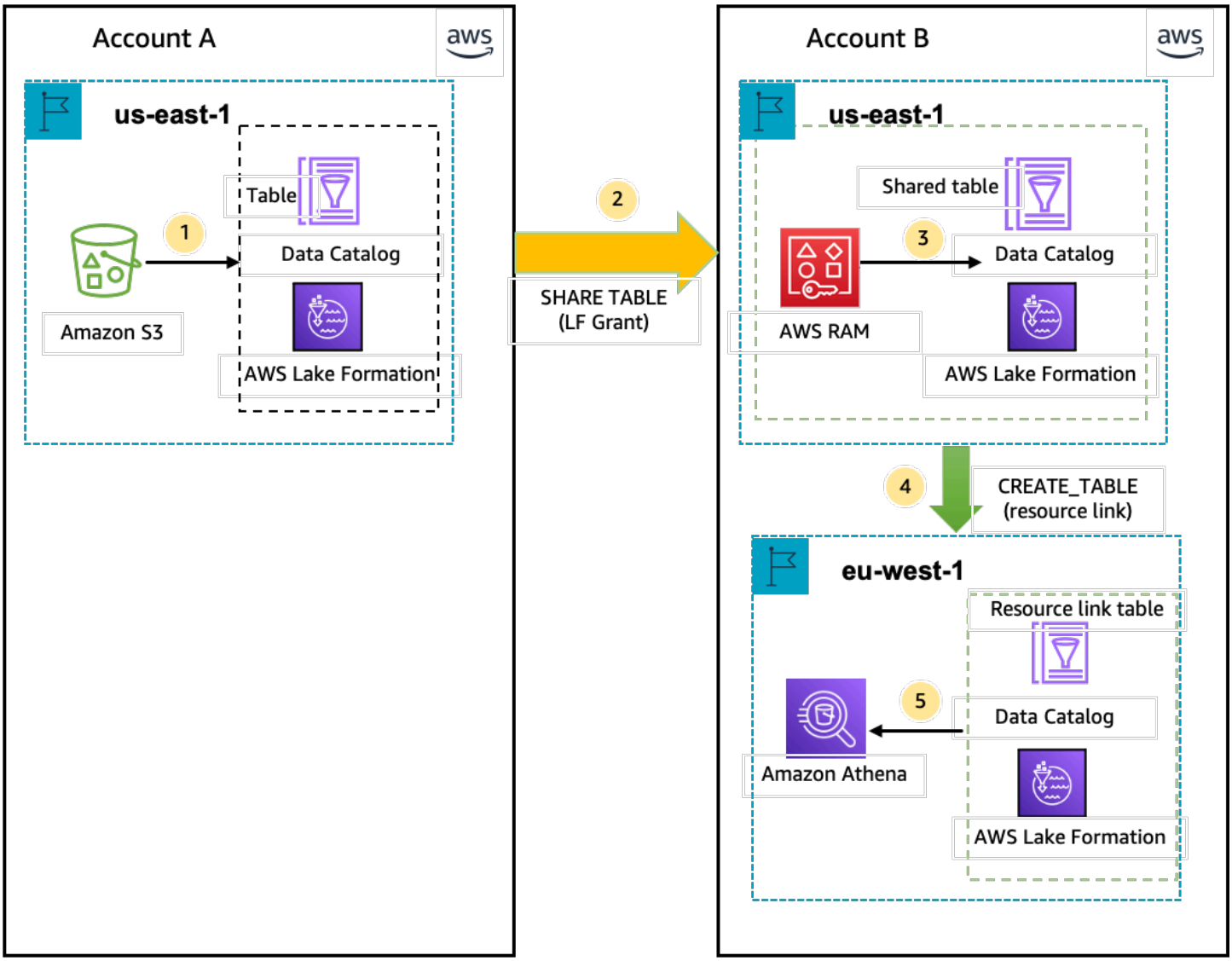
L'amministratore del data lake esegue le seguenti attività (passaggi 1-2):

1. Un amministratore del data lake configura un AWS account con i database e le tabelle Data Catalog e registra una posizione dati Amazon S3 con Lake Formation nella regione Stati Uniti orientali (Virginia settentrionale).

   Concede `Select` l'autorizzazione per una risorsa del Data Catalog (tabella dei prodotti nel diagramma) a un principale (utente) nello stesso account.
2. Crea un collegamento a una risorsa nella regione Europa (Irlanda) che punta alla tabella di origine nella regione Stati Uniti orientali (Virginia settentrionale). Concede l'`DESCRIBE`autorizzazione per il collegamento alla risorsa dalla regione Europa (Irlanda) al principale.
3. L'utente esegue una query sulla tabella dalla regione Europa (Irlanda) utilizzando Athena.

Flusso di lavoro per accedere alle tabelle condivise con un account esterno AWS

Nel diagramma seguente, l'account produttore (Account A) ospita il bucket Amazon S3, registra la posizione dei dati e condivide una tabella del catalogo dati con un account consumer (Account B) nella regione Stati Uniti orientali (Virginia settentrionale) e un utente dell'account consumer (Account B) esegue una query sulla tabella dalla regione Europa (Irlanda).



1. Un amministratore del data lake configura un AWS account (account produttore) con le risorse Data Catalog e una data location Amazon S3 registrata presso Lake Formation nella regione Stati Uniti orientali (Virginia settentrionale).
2. L'amministratore del data lake dell'account produttore condivide una tabella Data Catalog con un account consumatore.
3. L'amministratore del data lake dell'account consumer accetta l'invito alla condivisione dei dati nella regione Stati Uniti orientali (Virginia settentrionale) e concede l'`Select`autorizzazione per la tabella condivisa a un principale della stessa regione.
4. L'amministratore del data lake dell'account consumer crea un collegamento di risorse nella regione Europa (Irlanda) che rimanda alla tabella condivisa di destinazione nella regione Stati Uniti orientali

(Virginia settentrionale) e concede all'utente l'`DESCRIBE`autorizzazione per il collegamento alla risorsa dalla regione Europa (Irlanda).

5. L'utente interroga i dati dalla regione Europa (Irlanda) utilizzando Athena.

Configurazione dell'accesso alle tabelle tra regioni

Per accedere ai dati da una regione diversa, devi prima configurare i database e le tabelle del catalogo dati nella regione in cui registri la tua posizione dati su Amazon S3. Puoi condividere i database e le tabelle del Data Catalog con i principali del tuo account o di un altro account. Quindi, devi creare amministratori di data lake che possano creare collegamenti a risorse che puntino alla posizione dei dati condivisi di destinazione nelle regioni in cui gli utenti interrogano i dati.

Per interrogare i dati condivisi all'interno dello stesso account da una regione diversa

In questa sezione, la regione della tabella condivisa di destinazione viene denominata Regione A e gli utenti eseguono query dalla Regione B.

1. Configurazione dell'account nella Regione A (dove si creano e si condividono i dati)

   Un amministratore del data lake deve completare le seguenti azioni:

   a. Registra una posizione dati Amazon S3.

      Per ulteriori informazioni, consulta Aggiungere una posizione Amazon S3 al tuo data lake.

   b. Crea database e tabelle nell'account. Questa operazione può essere eseguita anche da un utente non amministrativo che dispone delle autorizzazioni per creare database e tabelle.

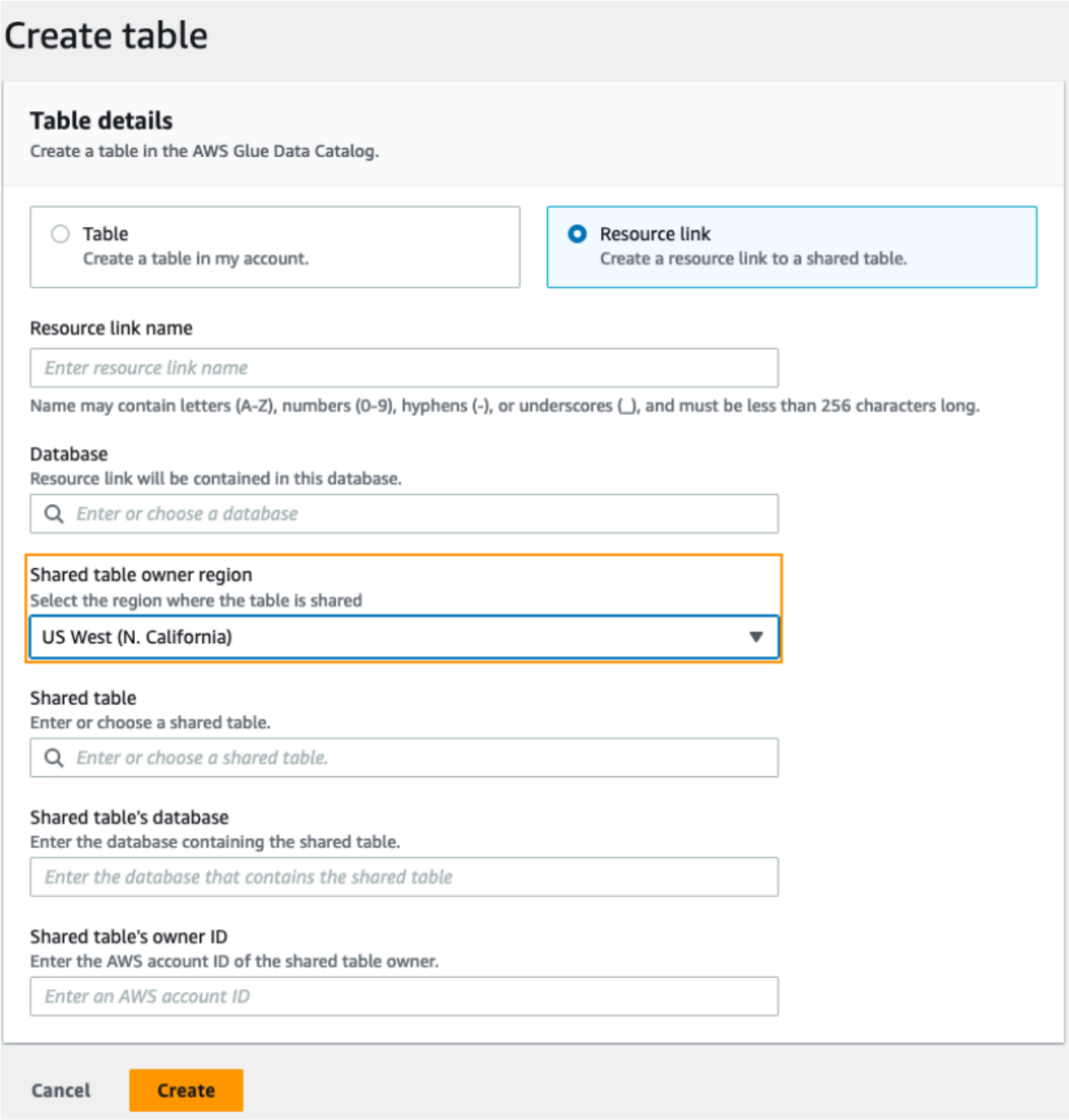
   c. Concedi le autorizzazioni relative ai dati su una tabella ai principali con. `Grantable permissions`

      Per ulteriori informazioni, consultare Concessione e revoca delle autorizzazioni per le risorse del Data Catalog.

2. Configurazione dell'account nella Regione B (dove si accede ai dati)

   Un amministratore del data lake deve completare le seguenti azioni:

   a. Crea un link di risorsa nella Regione B che punti alla tabella condivisa di destinazione nella Regione A. Specificare la regione proprietaria della tabella condivisa nella schermata Crea tabella.



Per istruzioni sulla creazione di collegamenti di risorse a database e tabelle, vedereCreazione di collegamenti alle risorse.

b. Concedi `Describe` l'autorizzazione ai responsabili IAM sul link alla risorsa nella Regione B.

Per ulteriori informazioni sulla concessione delle autorizzazioni sui link alle risorse, consulta. Concessione delle autorizzazioni per i collegamenti alle risorse

I responsabili IAM nella regione B possono interrogare la tabella di destinazione tramite il collegamento utilizzando Athena.

Per accedere ai dati di più account da una regione diversa

1. Configurazione dell'account produttore/concedente

   Un amministratore del data lake deve completare le seguenti azioni:

   a. Configura l'account produttore/concedente nella Regione A.
   b. Registra una posizione dati Amazon S3 nella regione A.
   c. Crea database e tabelle. Questa operazione può essere eseguita da un utente non amministrativo che dispone delle autorizzazioni necessarie per creare tabelle.
   d. Concedi le autorizzazioni relative ai dati all'account consumatore/beneficiario su una tabella nella Regione A con. `Grantable permissions`

      Per ulteriori informazioni, consulta Condivisione delle tabelle e dei database del Data Catalog tra i nostri principali IAM provenienti Account AWS da account esterni.
2. Configurazione dell'account consumatore/beneficiario

   Un amministratore del data lake deve completare le seguenti azioni:

   a. Accetta l'invito alla condivisione delle risorse dalla AWS RAM Regione A.
   b. Crea un link alla risorsa nella Regione B che punti alla tabella condivisa. La regione B è dove gli utenti vorranno interrogare la tabella.
   c. Concedi le autorizzazioni relative ai dati sulla tabella condivisa ai responsabili IAM nella regione A.

      > **Note**
      >
      > È necessario concedere le autorizzazioni alla tabella condivisa nella stessa regione in cui la tabella è stata condivisa.

   d. Concedi le autorizzazioni ai responsabili sul link alla risorsa nella Regione B.

      I responsabili dell'account consumatore nella regione B interrogano quindi la tabella condivisa dalla regione B utilizzando Athena.

Condivisione dei dati in AWS Lake Formation

Puoi utilizzare la funzionalità di condivisione AWS Lake Formation dei dati per concedere e gestire le autorizzazioni sui dati archiviati in posizioni diverse da Amazon S3 e sui metadati archiviati in posizioni diverse da. AWS Glue Data Catalog Con la funzionalità di condivisione dei dati, puoi configurare e gestire le autorizzazioni sui set di dati in Amazon Redshift senza migrare i dati in Amazon S3. Puoi anche utilizzare la funzionalità di federazione del catalogo dati per connetterti a metastore esterni.

Successivamente, puoi utilizzare Lake Formation per gestire i dati e le autorizzazioni di accesso in un Data Catalog centrale definendo politiche di controllo degli accessi granulari. Gli amministratori di Data Lake possono concedere le autorizzazioni ad altri responsabili IAM all'interno dell'account o su più account sulle risorse del Data Catalog. I responsabili IAM possono interrogare i dati condivisi utilizzando Amazon Redshift Spectrum e Amazon Athena.

Lake Formation offre i seguenti metodi per condividere dati e gestire le autorizzazioni su set di dati esterni e metastore esterni:

- Integrazione di Lake Formation con la condivisione dei dati di Amazon Redshift: utilizza Lake Formation per gestire centralmente le autorizzazioni di accesso a livello di database, tabelle, colonne e righe delle condivisioni di dati Amazon Redshift e limitare l'accesso degli utenti agli oggetti all'interno di un datashare.
- Connessione AWS Glue Data Catalog a metastore esterni: collega i metastore esterni AWS Glue Data Catalog per gestire le autorizzazioni di accesso ai set di dati in Amazon S3 utilizzando Lake Formation. Non è necessaria alcuna migrazione dei metadati in. AWS Glue Data Catalog
- Integrazione di Lake Formation con AWS Data Exchange — Lake Formation supporta la concessione di licenze di accesso ai dati tramite. AWS Data Exchange Se sei interessato a concedere in licenza i tuoi dati di Lake Formation, consulta Cosa c'è AWS Data Exchange nella Guida per l'AWS Data Exchangeutente.

Argomenti

Gestione delle autorizzazioni per i dati in un datashare Amazon Redshift

ConAWS Lake Formation, puoi gestire i dati in modo sicuro in un datashare di Amazon Redshift. Amazon Redshift è un servizio di data warehouse completamente gestito su scala di petabyte nel cloud. AWS Utilizzando la funzionalità di condivisione dei dati, Amazon Redshift ti aiuta a condividere i dati tra di loro. Account AWS Per ulteriori informazioni sulla condivisione dei dati di Amazon Redshift, consulta Panoramica sulla condivisione dei dati in Amazon Redshift.

In Amazon Redshift, l'amministratore del cluster di produttori crea un datashare e lo condivide con l'amministratore del data lake. Per step-by-step istruzioni sulla creazione di un amministratore di data lake, consulta. Crea un amministratore del data lake

Dopo che l'amministratore del data lake ha accettato il datashare, è necessario creare un AWS Glue Data Catalog database per il datashare specifico. In questo modo puoi controllarne l'accesso utilizzando le autorizzazioni di Lake Formation. Lake Formation mappa ogni datashare su un database Data Catalog corrispondente. Questi vengono visualizzati come database federati nel Data Catalog.

Un database viene definito database federato quando punta a un'entità esterna al Data Catalog. Le tabelle e le viste nel datashare Amazon Redshift sono elencate come singole tabelle nel Data Catalog. Puoi condividere il database federato con i principali IAM e gli utenti SAML selezionati all'interno dello stesso account o in un altro account con Lake Formation. Puoi anche includere espressioni di filtro per righe e colonne per limitare l'accesso a determinati dati. Per ulteriori informazioni, consulta Panoramica del filtraggio dei dati.

Per fornire agli utenti l'accesso a un datashare Amazon Redshift, devi fare quanto segue:

1. Aggiorna le impostazioni del Data Catalog per abilitare le autorizzazioni di Lake Formation.
2. Accetta l'invito al datashare dall'amministratore del cluster di produttori di Amazon Redshift e registra il datashare in Lake Formation.

   Dopo aver completato questo passaggio, puoi gestire il datashare all'interno del Lake Formation Data Catalog.
3. Crea un database federato e definisci le autorizzazioni su quel database.
4. Concedi le autorizzazioni agli utenti su database e tabelle. Puoi condividere l'intero database o un sottoinsieme di tabelle con gli utenti dello stesso account o di un altro account.

Per le limitazioni, consulta Limitazioni della condivisione dei dati di Amazon Redshift.

Argomenti

Prerequisiti per la configurazione delle autorizzazioni sulle condivisioni di dati Amazon Redshift

Aggiorna le impostazioni predefinite del Data Catalog

Per abilitare le autorizzazioni Lake Formation per le risorse Data Catalog, ti consigliamo di disabilitare le impostazioni predefinite del Data Catalog in Lake Formation. Per ulteriori informazioni, consulta Modifica il modello di autorizzazione predefinito o utilizza la modalità di accesso ibrida.

Aggiorna le autorizzazioni

Oltre alle autorizzazioni di amministratore del data lake (`AWSLakeFormationDataAdmin`), sono necessarie anche le seguenti autorizzazioni per accettare un datashare Amazon Redshift in Lake Formation:

- `glue:PassConnection on aws:redshift`
- `redshift:AssociateDataShareConsumer`
- `redshift:DescribeDataSharesForConsumer`
- `redshift:DescribeDataShares`

L'utente IAM dell'amministratore del data lake dispone implicitamente delle seguenti autorizzazioni.

- data_location_access
- creare_database
- Lake Formation: registra la risorsa

Configurazione delle autorizzazioni per le condivisioni di dati Amazon Redshift

Questo argomento descrive i passaggi da seguire per accettare un invito alla condivisione di dati, creare un database federato e concedere le autorizzazioni. Puoi usare la console Lake Formation o il AWS Command Line Interface (AWS CLI). Gli esempi in questo argomento mostrano il cluster di produttori, il Data Catalog e il consumatore di dati nello stesso account.

Per ulteriori informazioni sulle funzionalità cross-account di Lake Formation, consultaCondivisione dei dati tra account in Lake Formation.

Per configurare le autorizzazioni per un datashare

1. Rivedi un invito al datashare e accettalo.

   Console

   1. Accedi alla console di Lake Formation come amministratore del data lake all'indirizzo https://console.aws.amazon.com/lakeformation/. Vai alla pagina di condivisione dei dati.
   2. Controlla le condivisioni di dati a cui sei autorizzato ad accedere. La colonna Stato indica il tuo attuale stato di partecipazione al datashare. Lo stato In sospeso indica che sei stato aggiunto a un datashare, ma non l'hai ancora accettato o hai rifiutato l'invito.
   3. Per rispondere a un invito al datashare, seleziona il nome del datashare e scegli Rivedi l'invito. In Accetta o rifiuta il datashare, esamina i dettagli dell'invito. Scegli Accetta per accettare l'invito o Rifiuta per rifiutarlo. Non avrai accesso al datashare se rifiuti l'invito.

   AWS CLI

   Gli esempi seguenti mostrano come visualizzare, accettare e registrare l'invito. Sostituisci l'Account AWSID con un Account AWS ID valido. Sostituisci `data-share-arn` con l'effettivo Amazon Resource Name (ARN) che fa riferimento al datashare.

   1. Visualizza un invito in sospeso.

      ```
      aws redshift describe-data-shares \
       --data-share-arn 'arn:aws:redshift:us-
      east-1:111122223333:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
      federatedds' \
      ```

2. Accetta un datashare.

```
 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. Registra il datashare nell'account Lake Formation. Utilizza l'operazione RegisterResourceAPI per registrare il datashare in Lake Formation. `DataShareArn`è il parametro di input per. `ResourceArn`

> **Note**
>
> Si tratta di un passaggio obbligatorio.

```
aws lakeformation register-resource \
 --resource-arn 'arn:aws:redshift:us-
east-1:111122223333:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
federatedds'
```

2. Crea un database.

   Dopo aver accettato un invito alla condivisione di dati, devi creare un database che punti al database Amazon Redshift associato al datashare. Devi essere un amministratore di data lake per creare un database.

   Console

   1. Seleziona il datashare dal riquadro Inviti e scegli Imposta i dettagli del database.
   2. In Imposta i dettagli del database, inserisci un nome e un identificatore univoci per il datashare. Questo identificatore viene utilizzato per mappare il datashare internamente nella gerarchia dei metadati (dbname.schema.table).
   3. Scegli Avanti per concedere le autorizzazioni ad altri utenti sul database e sulle tabelle condivisi.

AWS CLI

Utilizza il codice di esempio seguente per creare un database che punti al database Amazon Redshift condiviso con Lake Formation utilizzando il. 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. Concedi le autorizzazioni.

Dopo aver creato il database, puoi concedere le autorizzazioni agli utenti del tuo account o a organizzazioni esterneAccount AWS. Non potrai concedere autorizzazioni per i dati di scrittura (inserimento, eliminazione) e per i metadati (modifica, eliminazione, creazione) sul database federato mappato su un datashare Amazon Redshift. Per ulteriori informazioni sulla concessione delle autorizzazioni, consulta. Gestione delle autorizzazioni di Lake Formation

> **Note**
>
> In qualità di amministratore del data lake, puoi visualizzare solo le tabelle nei database federati. Per eseguire qualsiasi altra azione, è necessario concedere ulteriori autorizzazioni su tali tabelle.

Console

1. Nella schermata Concedi autorizzazioni, seleziona gli utenti a cui concedere le autorizzazioni.
2. Scegli Concessione.

AWS CLI

Utilizza gli esempi seguenti per concedere le autorizzazioni per database e tabelle utilizzando: 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"
                  ]
}
```

Interrogazione di database federati

Dopo aver concesso le autorizzazioni, gli utenti possono accedere e iniziare a interrogare il database federato utilizzando Amazon Redshift. Gli utenti possono ora utilizzare il nome del database locale per fare riferimento al datashare Amazon Redshift nelle query SQL. In Amazon Redshift, la tabella clienti nello schema pubblico condiviso tramite il datashare avrà una tabella corrispondente creata come `public.customer` nel Data Catalog.

1. Prima di interrogare il database federato utilizzando Amazon Redshift, l'amministratore del cluster crea un database dal database Data Catalog utilizzando il seguente comando:

```
CREATE DATABASE sharedcustomerdb FROM ARN
 'arn:aws:glue:<region>:111122223333:database/tahoedb' WITH DATA CATALOG SCHEMA
 tahoedb
```

2. L'amministratore del cluster concede le autorizzazioni di utilizzo sul database.

```
GRANT USAGE ON DATABASE sharedcustomerdb TO IAM:user;
```

3. L'utente federato può ora accedere agli strumenti SQL per interrogare la tabella.

```
Select * from sharedcustomerdb.public.customer limit 10;
```

Per ulteriori informazioni, consulta la sezione Querying AWS Glue Data Catalog in Amazon Redshift Management Guide.

Gestione delle autorizzazioni sui set di dati che utilizzano metastore esterni

Con la federazione dei AWS Glue Data Catalog metadati (Data Catalog federation), puoi connettere il Data Catalog a metastore esterni che archiviano i metadati per i tuoi dati Amazon S3 e gestire in modo sicuro le autorizzazioni di accesso ai dati utilizzando. AWS Lake Formation Non è necessario migrare i metadati dal metastore esterno al Data Catalog.

Il Data Catalog fornisce un archivio centralizzato di metadati che semplifica la gestione e la scoperta dei dati su sistemi diversi. Quando la tua organizzazione gestisce i dati nel catalogo dati, puoi utilizzarli AWS Lake Formation per controllare l'accesso ai tuoi set di dati in Amazon S3.

> **Note**
>
> Attualmente, supportiamo solo la federazione dei metastore di Apache Hive (versione 3 e successive).

Per configurare la federazione di Data Catalog, forniamo un'applicazione AWS Serverless Application Model (AWS SAM) chiamata GlueDataCatalogFederation - in. HiveMetastore AWS Serverless Application Repository

L'implementazione di riferimento è fornita GitHub come progetto open source presso AWS Glue Data CatalogFederation - Hive Metastore.

L'AWS SAMapplicazione crea e distribuisce le seguenti risorse necessarie per connettere il Data Catalog al metastore Hive:

- Una AWS Lambda funzione: ospita l'implementazione del servizio federativo che comunica tra il Data Catalog e il metastore Hive. AWS Gluerichiama questa funzione Lambda per recuperare oggetti di metadati dal metastore Hive.
- Amazon API Gateway— L'endpoint di connessione per il metastore Hive che funge da proxy per indirizzare tutte le chiamate alla funzione Lambda.
- Un ruolo IAM: un ruolo con le autorizzazioni necessarie per creare la connessione tra il Data Catalog e il metastore Hive.
- AWS Glueconnessione: un Amazon API Gateway tipo di AWS Glue connessione che memorizza l'Amazon API Gatewayendpoint e un ruolo IAM per richiamarlo.

Quando esegui una query sulle tabelle, il AWS Glue servizio effettua una chiamata di runtime al metastore Hive e recupera i metadati. La funzione Lambda funge da traduttore tra il metastore Hive e Data Catalog.

Dopo aver stabilito la connessione, per sincronizzare i metadati nel metastore Hive con il Data Catalog, è necessario creare un database federato nel Data Catalog utilizzando i dettagli di connessione del metastore Hive e mappare questo database al database Hive. Un database viene definito database federato quando punta a un'entità esterna al Data Catalog.

Puoi applicare le autorizzazioni di Lake Formation utilizzando il controllo degli accessi basato su tag e il metodo della risorsa denominata sul database federato e condividerlo tra più Account AWS unità organizzative (OU). AWS Organizations Puoi anche condividere il database federato direttamente con i responsabili IAM di un altro account.

Puoi definire autorizzazioni dettagliate a livello di colonna, riga e cella utilizzando i filtri dati di Lake Formation sulle tabelle Hive esterne. Puoi usare Amazon Athena, Amazon Redshift o Amazon EMR per interrogare le tabelle Hive esterne gestite da Lake Formation.

Per ulteriori informazioni sulla condivisione e il filtraggio dei dati tra account, consulta:

- Condivisione dei dati tra account in Lake Formation
- Filtraggio dei dati e sicurezza a livello di cella in Lake Formation

Fasi di alto livello per la federazione dei metadati di Data Catalog

1. Crei utenti e ruoli IAM con le autorizzazioni appropriate per distribuire l'AWS SAMapplicazione e creare database federati.
2. Registri la posizione dei dati di Amazon S3 con Lake Formation selezionando l'`Enable Data Catalog federation`opzione per i set di dati che utilizzano un metastore Hive esterno.
3. È possibile configurare le impostazioni AWS SAM dell'applicazione (nome della AWS Glue connessione, URL del metastore Hive e parametri della funzione Lambda) e distribuire l'applicazione. AWS SAM
4. L'AWS SAMapplicazione distribuisce le risorse necessarie per connettere il metastore Hive esterno al Data Catalog.
5. Per applicare le autorizzazioni di Lake Formation al database e alle tabelle Hive, crei un database nel Data Catalog utilizzando i dettagli di connessione del metastore Hive e mappi questo database al database Hive.

6. Concedi le autorizzazioni sui database federati ai responsabili del tuo account o di un altro account.

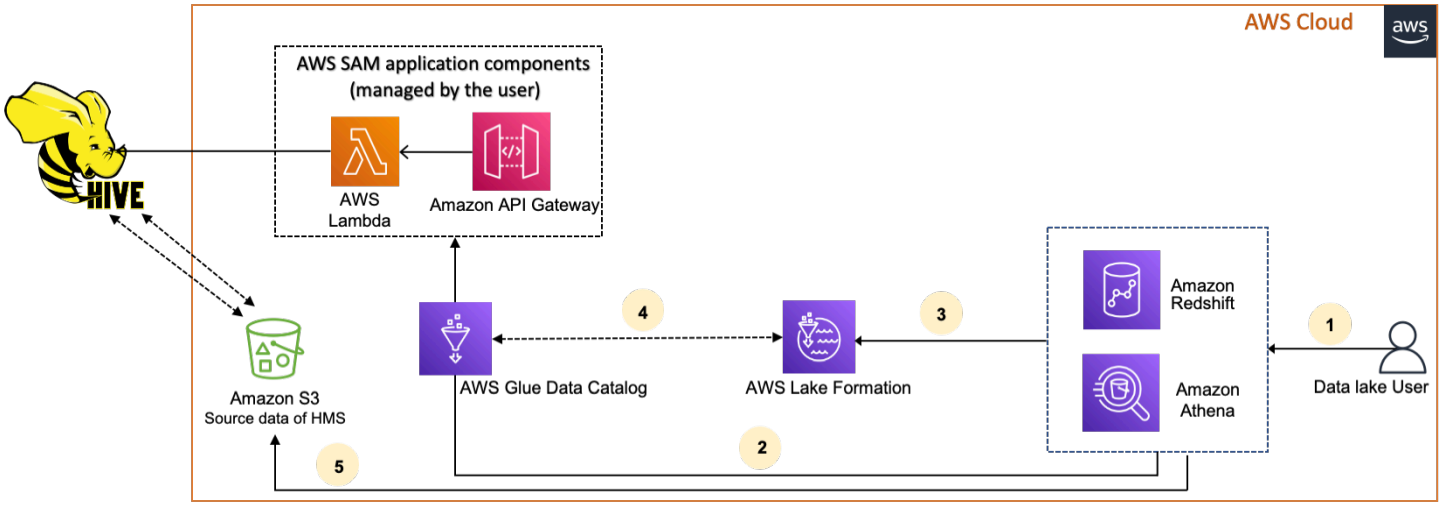
> **Note**
>
> Puoi connettere il Data Catalog a un mestastore Hive esterno, creare database federati ed eseguire query e script ETL su database e tabelle Hive senza applicare le autorizzazioni di Lake Formation. Per i dati di origine in Amazon S3 che non sono registrati con Lake Formation, l'accesso è determinato dalle policy di autorizzazione IAM per Amazon S3 e dalle operazioni AWS Glue.

Per le limitazioni, consulta Considerazioni e limitazioni sulla condivisione dei dati dei metadati di Hive.

Argomenti

Flusso di lavoro

Il diagramma seguente mostra il flusso di lavoro per il collegamento di un metastore Hive esterno. AWS Glue Data Catalog



1. Un principale invia una richiesta utilizzando un servizio integrato come Athena o Redshift Spectrum.
2. Il servizio integrato effettua una chiamata al Data Catalog per i metadati, che a sua volta richiama l'endpoint Hive metastore disponibile e riceve risposte alle Amazon API Gateway richieste di metadati.
3. Il servizio integrato invia la richiesta a Lake Formation per verificare le informazioni sulla tabella e le credenziali per accedere alla tabella.
4. Lake Formation autorizza la richiesta e invia credenziali temporanee all'applicazione integrata, che consente l'accesso ai dati.
5. Utilizzando le credenziali temporanee ricevute da Lake Formation, il servizio integrato legge i dati da Amazon S3 e condivide i risultati con il responsabile.

Prerequisiti per connettere il Data Catalog al metastore Hive

Per connetterlo AWS Glue Data Catalog a un metastore Apache Hive esterno e configurare le autorizzazioni di accesso ai dati, è necessario soddisfare i seguenti requisiti:

> **Note**
>
> Consigliamo che un amministratore di Lake Formation distribuisca l'AWS SAMapplicazione e solo un utente privilegiato utilizzi la connessione metastore Hive per creare i database federati corrispondenti.

1. Creare ruoli IAM.

   Per distribuire l'applicazione AWS SAM

   - Crea un ruolo con le autorizzazioni necessarie per la distribuzione delle risorse (funzione LambdaAmazon API Gateway, ruolo IAM e AWS Glue connessione) necessarie per creare una connessione al metastore Hive.

   Per creare database federati

   Le seguenti autorizzazioni sono richieste per le risorse:

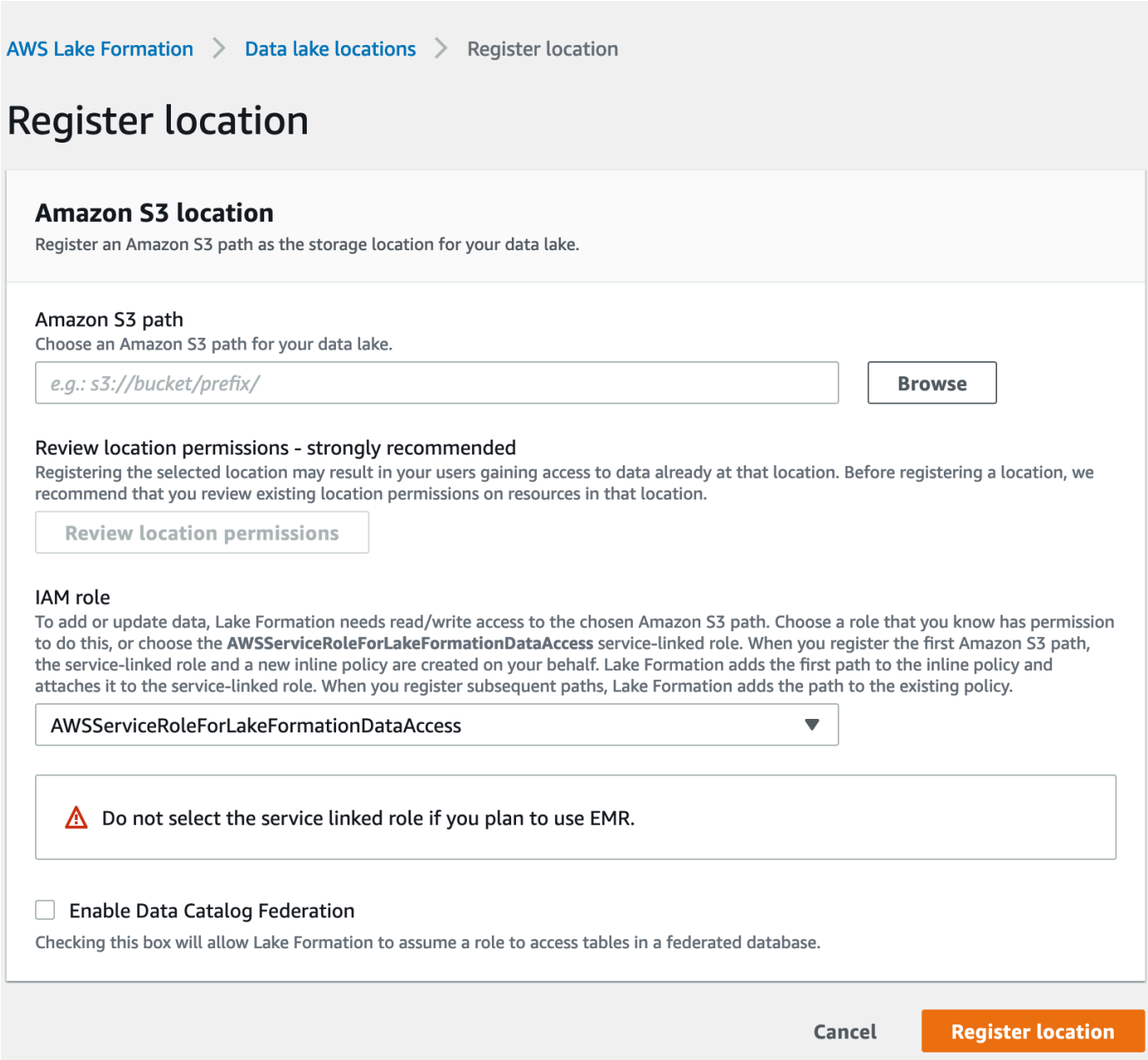
- `glue:CreateDatabase on resource arn:aws:glue:region:account-id:database/gluedatabasename`
- `glue:PassConnection on resource arn:aws:glue:region:account-id:connection/hms_connection`

2. Registra la sede Amazon S3 con Lake Formation.

   Per utilizzare Lake Formation per gestire e proteggere i dati nel tuo data lake, devi registrare la posizione Amazon S3 che contiene i dati per le tabelle nel metastore Hive con Lake Formation. In questo modo, Lake Formation può fornire credenziali per servizi di AWS analisi come Athena, Redshift Spectrum e Amazon EMR.

   Per ulteriori informazioni sulla registrazione di una sede Amazon S3, consulta. Aggiungere una posizione Amazon S3 al tuo data lake

   Quando registri la posizione Amazon S3, seleziona la casella di controllo Enable Data Catalog Federation per consentire a Lake Formation di assumere un ruolo per accedere alle tabelle in un database federato.



Per ulteriori informazioni sulla registrazione di una posizione dati con Lake Formation, vedereConfigura una posizione Amazon S3 per il tuo data lake.

3. Usa la versione corretta di Amazon EMR.

   Per utilizzare Amazon EMR con i database metastore Hive federati, devi disporre della versione 3.x o successiva di Hive e della versione 6.x o successiva di Amazon EMR.

Connessione del Data Catalog a un metastore Hive esterno

Per connetterlo AWS Glue Data Catalog a un metastore Hive, devi distribuire un'applicazione chiamata -. AWS SAM GlueDataCatalogFederation HiveMetastore Crea le risorse necessarie per connettere il metastore Hive esterno al Data Catalog. È possibile accedere all'AWS SAMapplicazione in. AWS Serverless Application Repository

L'AWS SAMapplicazione crea la connessione per il metastore Hive dietro Amazon API Gateway utilizzando una funzione Lambda. L'AWS SAMapplicazione utilizza un URI (Uniform Resource Identifier) come input dell'utente e collega il metastore Hive esterno al Data Catalog. Quando un utente esegue una query sulle tabelle Hive, il Data Catalog chiama l'endpoint API Gateway. L'endpoint richiama la funzione Lambda per recuperare i metadati delle tabelle Hive.

Per connettere il Data Catalog al metastore Hive e configurare le autorizzazioni

1. Distribuisci l'applicazione. AWS SAM
    1. Accedi alla AWS Management Console e apri la AWS Serverless Application Repository.
    2. Nel pannello di navigazione, scegli Available applications (Applicazioni disponibili).
    3. Scegli Applicazioni pubbliche.
    4. Seleziona l'opzione Visualizzare le app che creano ruoli IAM personalizzati o policy delle risorse.
    5. Nella casella di ricerca, inserisci il nome GlueDataCatalogFederation- HiveMetastore.
    6. Scegli l'HiveMetastoreapplicazione GlueDataCatalogFederation-.
    7. In Impostazioni dell'applicazione, inserisci le seguenti impostazioni minime richieste per la tua funzione Lambda:
        - Nome dell'applicazione: un nome per l'AWS SAMapplicazione.
        - GlueConnectionName- Un nome per la connessione.
        - HiveMetastoreURI: l'URI del tuo host metastore Hive.
        - LambdaMemory- La quantità di memoria Lambda in MB è compresa tra 128 e 10240. Il valore di default è 1024.
        - LambdaTimeout- La durata massima di invocazione Lambda in secondi. Il valore predefinito è 30.
        - VPC e SecurityGroupIds VPCSubnetIds: informazioni per il VPC in cui è presente il metastore Hive.

8. Seleziona I acknowledge that this app creates custom IAM roles and resource policies (Sono consapevole che questa app crea ruoli IAM personalizzati e policy della risorsa). Per ulteriori informazioni, scegliere il link Info (Informazioni) .
9. Nella parte inferiore destra della pagina Application settings (Impostazioni dell'applicazione), scegli Deploy (Implementa). Al termine dell'implementazione, la funzione Lambda viene visualizzata nella sezione Resources (Risorse) nella console Lambda.

L'applicazione viene distribuita su Lambda. Il suo nome è preceduto da serverlessrepo- per indicare che l'applicazione è stata distribuita da. AWS Serverless Application Repository Selezionando l'applicazione si accede alla pagina Risorse in cui sono elencate tutte le risorse dell'applicazione che sono state distribuite. Le risorse includono la funzione Lambda che consente la comunicazione tra il Data Catalog e il metastore Hive, la AWS Glue connessione e altre risorse necessarie per la federazione del database.

2. Crea un database federato nel Data Catalog.

   Dopo aver creato una connessione al metastore Hive, puoi creare database federati nel Data Catalog che puntano ai database metastore Hive esterni. Devi creare un database corrispondente nel Data Catalog per ogni database di metastore Hive che stai connettendo al Data Catalog.

   Lake Formation console

   1. Nella pagina Condivisione dati, scegli la scheda Database condivisi, quindi scegli Crea database.
   2. Per Nome della connessione, scegli il nome della tua connessione al metastore Hive dal menu a discesa.
   3. Inserisci un nome di database univoco e l'identificatore di origine della federazione per il database. Questo è il nome che usi nelle istruzioni SQL quando esegui query sulle tabelle. Il nome può essere composto da un massimo di 255 caratteri e deve essere univoco all'interno del tuo account.
   4. Scegliere Crea database.

   AWS CLI

   ```
   aws glue create-database \
   '{
   ```

```
 "CatalogId": "<111122223333>",
  "database-input": {
    "Name":"<fed_glue_db>",
    "FederatedDatabase":{
        "Identifier":"<hive_db_on_emr>",
        "ConnectionName":"<hms_connection>"
     }
   }
 }'
```

3. Visualizza le tabelle nel database federato.

   Dopo aver creato il database federato, puoi visualizzare l'elenco delle tabelle nel tuo metastore Hive utilizzando la console Lake Formation o il. AWS CLI

   Lake Formation console

   1. Seleziona il nome del database dalla scheda Database condivisi.
   2. Nella pagina Database, scegli Visualizza tabelle.

   AWS CLI

   Gli esempi seguenti mostrano come recuperare la definizione di connessione, il nome del database e alcune o tutte le tabelle del database. Sostituisci l'ID del catalogo dati con l'Account AWSID valido utilizzato per creare il database. Sostituisci `hms_connection` con il nome della connessione.

   ```
   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. Concedi le autorizzazioni.

   Dopo aver creato il database, puoi concedere le autorizzazioni ad altri utenti e ruoli IAM nel tuo account o a organizzazioni esterneAccount AWS. Non sarai in grado di concedere autorizzazioni per i dati di scrittura (inserimento, eliminazione) e autorizzazioni per i metadati (modifica, eliminazione, creazione) sui database federati. Per ulteriori informazioni sulla concessione delle autorizzazioni, vedere. Gestione delle autorizzazioni di Lake Formation

5. Interroga i database federati.

   Dopo aver concesso le autorizzazioni, gli utenti possono accedere e iniziare a interrogare il database federato utilizzando Athena e Amazon Redshift. Gli utenti possono ora utilizzare il nome del database locale per fare riferimento al database Hive nelle query SQL.

   Sintassi di interrogazione di esempio Amazon Athena

   Sostituisci `fed_glue_db` con il nome del database locale creato in precedenza.

   ```
   Select * from fed_glue_db.customers limit 10;
   ```

Risorse aggiuntive

Il seguente post sul blog contiene istruzioni dettagliate per configurare i permessi di Lake Formation su un database e tabelle di metastore Hive e interrogarli utilizzando Athena. Illustriamo anche un caso d'uso della condivisione tra account, in cui un responsabile di Lake Formation nell'account produttore A condivide un database Hive federato e tabelle che utilizzano il tag LF con l'account consumatore B.

- Interroga il tuo metastore Apache Hive con le autorizzazioni AWS Lake Formation

Sicurezza in AWS Lake Formation

Per AWS, la sicurezza del cloud ha la massima priorità. In quanto cliente AWS, è possibile trarre vantaggio da un'architettura di data center e di rete progettata per soddisfare i requisiti delle organizzazioni più esigenti a livello di sicurezza.

La sicurezza è una responsabilità condivisa tra te e AWS. Il modello di responsabilità condivisa descrive questo come sicurezza del cloud e sicurezza nel cloud:

- La sicurezza del cloud: AWS è responsabile della protezione dell'infrastruttura che esegue AWS i servizi nel cloud AWS. AWS fornisce, inoltre, servizi utilizzabili in modo sicuro. I revisori di terze parti testano e verificano regolarmente l'efficacia della sicurezza come parte dei programmi di conformità AWS. Per ulteriori informazioni sui programmi di conformità che si applicano a AWS Lake Formation, consulta Servizi coperti dal programma di conformitàAWS.
- Sicurezza nel cloud: la tua responsabilità è determinata dal servizio AWS che utilizzi. L'utente è anche responsabile per altri fattori, tra cui la riservatezza dei dati, i requisiti dell'azienda, le leggi e le normative applicabili.

Questa documentazione consente di comprendere come applicare il modello di responsabilità condivisa quando utilizzisi usa Lake Formation. I seguenti argomenti illustrano come configurare Lake Formation per soddisfare gli obiettivi di sicurezza e conformità. È inoltre illustrato come utilizzare altri AWS servizi che consentono di monitorare e proteggere le risorse di Lake Formation.

Argomenti

- Protezione dei dati in Lake Formation
- Sicurezza dell'infrastruttura in AWS Lake Formation
- Prevenzione del confused deputy tra servizi
- Registrazione degli eventi di sicurezza AWS Lake Formation

Protezione dei dati in Lake Formation

Il modello di responsabilità AWS condivisa modello di di si applica alla protezione dei dati in AWS Lake Formation. Come descritto in questo modello, AWS è responsabile della protezione dell'infrastruttura globale che esegue tutto l'Cloud AWS. L'utente è responsabile del controllo dei contenuti ospitati su questa infrastruttura. Inoltre, sei responsabile della configurazione della

protezione e delle attività di gestione per i Servizi AWS che utilizzi. Per ulteriori informazioni sulla privacy dei dati, vedi le Domande frequenti sulla privacy dei dati. Per informazioni sulla protezione dei dati in Europa, consulta il post del blog relativo al Modello di responsabilità condivisa AWS e GDPR nel Blog sulla sicurezza AWS.

Per garantire la protezione dei dati, ti suggeriamo di proteggere le credenziali Account AWS e di configurare singoli utenti con AWS IAM Identity Center o AWS Identity and Access Management (IAM). In tal modo, a ogni utente verranno assegnate solo le autorizzazioni necessarie per svolgere i suoi compiti. Ti suggeriamo, inoltre, di proteggere i dati nei seguenti modi:

- Utilizza l'autenticazione a più fattori (MFA) con ogni account.
- Utilizza SSL/TLS per comunicare con le risorse AWS. È richiesto TLS 1.2 ed è consigliato TLS 1.3.
- Configura l'API e la registrazione delle attività degli utenti con AWS CloudTrail.
- Utilizza le soluzioni di crittografia AWS, insieme a tutti i controlli di sicurezza predefiniti in Servizi AWS.
- Utilizza i servizi di sicurezza gestiti avanzati, come Amazon Macie, che aiutano a individuare e proteggere i dati sensibili archiviati in Amazon S3.
- Se necessiti di moduli crittografici convalidati FIPS 140-2 quando accedi ad AWSattraverso un'interfaccia a riga di comando o un'API, utilizza un endpoint FIPS. Per ulteriori informazioni sugli endpoint FIPS disponibili, consulta il Federal Information Processing Standard (FIPS) 140-2.

Ti consigliamo vivamente di non inserire mai informazioni riservate o sensibili, ad esempio gli indirizzi e-mail dei clienti, nei tag o nei campi di testo in formato libero, ad esempio nel campo Nome. Ciò include quando lavori con Lake Formation o altro Servizi AWS utilizzando la console, l'API o AWS gli SDK. AWS CLI I dati inseriti nei tag o nei campi di testo in formato libero utilizzati per i nomi possono essere utilizzati per i la fatturazione o i log di diagnostica. Quando fornisci un URL a un server esterno, ti suggeriamo vivamente di non includere informazioni sulle credenziali nell'URL per convalidare la tua richiesta al server.

Crittografia dei dati inattivi

AWS Lake Formationsupporta la crittografia dei dati nelle seguenti aree:

- Dati nel tuo data lake Amazon Simple Storage Service (Amazon S3).

  Lake Formation supporta la crittografia dei dati con AWS Key Management Service(AWS KMS). I dati vengono in genere scritti nel data lake tramite processi di AWS Glue estrazione,

trasformazione e caricamento (ETL). Per informazioni su come crittografare i dati scritti dai AWS Glue job, consulta Encrypting Data Written by Crawlers, Jobs and Development Endpoints nella Developer Guide. AWS Glue

- IIAWS Glue Data Catalog, dove Lake Formation archivia le tabelle di metadati che descrivono i dati nel data lake.

  Per ulteriori informazioni, consulta Encrypting Your Data Catalog nella AWS GlueDeveloper Guide.

Per aggiungere una posizione Amazon S3 come storage nel tuo data lake, devi registrare la posizione con. AWS Lake Formation Puoi quindi utilizzare le autorizzazioni di Lake Formation per un controllo granulare degli accessi agli AWS Glue Data Catalog oggetti che puntano a questa posizione e ai dati sottostanti nella posizione.

Lake Formation supporta la registrazione di una posizione Amazon S3 che contiene dati crittografati. Per ulteriori informazioni, consulta Registrazione di una posizione Amazon S3 crittografata.

Sicurezza dell'infrastruttura in AWS Lake Formation

In qualità di servizio gestito, AWS Lake Formation è protetto dalle procedure di sicurezza di rete globali di AWS descritte nel whitepaper Amazon Web Services: Panoramica dei processi di sicurezza.

Utilizza le chiamate all'APIAWS pubblicate da per accedere a Lake Formation tramite la rete. I client devono supportare Transport Layer Security (TLS) 1.0 o versioni successive. È consigliabile TLS 1.2 o versioni successive. I client devono, inoltre, supportare le suite di cifratura con PFS (Perfect Forward Secrecy), ad esempio Ephemeral Diffie-Hellman (DHE) o Elliptic Curve Ephemeral Diffie-Hellman (ECDHE). La maggior parte dei sistemi moderni, come Java 7 e versioni successive, supporta tali modalità.

Inoltre, le richieste devono essere firmate utilizzando un ID chiave di accesso e una chiave di accesso segreta associata a un principale IAM. In alternativa, è possibile utilizzare AWS Security Token Service (AWS STS) per generare le credenziali di sicurezza temporanee per sottoscrivere le richieste.

Prevenzione del confused deputy tra servizi

Con "confused deputy" si intende un problema di sicurezza in cui un'entità che non dispone dell’autorizzazione per eseguire una certa operazione può costringere un'entità con più privilegi a

eseguire tale operazione. In AWS, la rappresentazione cross-service può comportare il problema confused deputy. La rappresentazione tra servizi può verificarsi quando un servizio (il servizio chiamante) effettua una chiamata a un altro servizio (il servizio chiamato). Il servizio chiamante può essere manipolato per utilizzare le proprie autorizzazioni e agire sulle risorse di un altro cliente, a cui normalmente non avrebbe accesso. Per evitare ciò, AWS fornisce strumenti per poterti a proteggere i tuoi dati per tutti i servizi con entità di servizio a cui è stato concesso l'accesso alle risorse del tuo account.

Si consiglia di utilizzare le chiavi di contesto delle condizioni globali `aws:SourceArn` e `aws:SourceAccount` nelle policy delle risorse per limitare le autorizzazioni con cui AWS Lake Formation fornisce un altro servizio alla risorsa. Se si utilizzano entrambe le chiavi di contesto delle condizioni globali, il valore `aws:SourceAccount` e l'account nel valore `aws:SourceArn` devono utilizzare lo stesso ID account nella stessa istruzione di policy.

Attualmente, Lake Formation supporta solo`aws:SourceArn`Nel seguente formato:

```
arn:aws:lakeformation:aws-region:account-id:*
```

L'esempio seguente mostra il modo in cui puoi utilizzare il`aws:SourceArn`e`aws:SourceAccount`chiavi del contesto delle condizioni globali in Lake Formation per prevenire il problema confuso dei deputati.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "ConfusedDeputyPreventionExamplePolicy",
      "Effect": "Allow",
      "Principal": {
        "Service": "lakeformation.amazonaws.com"
      },
      "Action": [
        "sts:AssumeRole"
      ],
      "Condition": {
        "StringEquals": {
          "aws:SourceAccount": "account-id"
        },
        "ArnEquals": {
          "aws:SourceArn": "arn:aws:lakeformation:aws-region:account-id:*"
        }
```

```
      }
    }
  ]
}
```

Registrazione degli eventi di sicurezza AWS Lake Formation

AWSLake Formation è integrato conAWS CloudTrail, un servizio che fornisce una registrazione delle azioni intraprese da un utente, ruolo o AWS servizio in Lake Formation. CloudTrail acquisisce tutte le chiamate API per Lake Formation come eventi. Le chiamate acquisite includono chiamate dalla console di Lake FormationAWS Command Line Interface, e chiamate in codice alle operazioni dell'API Lake Formation.

Per ulteriori informazioni sulla registrazione degli eventi in Lake Formation, vedereRegistrazione di logAWSChiamate API Lake Formation utilizzandoAWS CloudTrail.

> **Note**
>
> `GetTableObjectsUpdateTableObjects`, e `GetWorkUnitResults` sono operazioni sul piano dati ad alto volume. Le chiamate a queste API non sono attualmente registrate. CloudTrail Per ulteriori informazioni sulle operazioni del piano dati in CloudTrail, consulta Registrazione degli eventi relativi ai dati per i percorsi nella Guida per l'AWS CloudTrailutente.
>
> Le modifiche apportate a Lake Formation per supportare CloudTrail eventi aggiuntivi saranno documentate suCronologia dei documenti per AWS Lake Formation.

Integrazione di servizi di terze parti con Lake Formation

L'integrazione con AWS Lake Formation consente ai servizi di terze parti di accedere in modo sicuro ai dati nei loro data lake basati su Amazon S3. Puoi utilizzare Lake Formation come motore di autorizzazione per gestire o applicare le autorizzazioni al tuo data lake con AWS servizi integrati come Amazon Athena, Amazon EMR e Redshift Spectrum. Lake Formation offre due opzioni per l'integrazione dei servizi:

1. Le impostazioni di integrazione delle applicazioni Lake Formation: Lake Formation può fornire credenziali temporanee ridotte sotto forma di token AWS STS alle sedi Amazon S3 registrate in base alle autorizzazioni effettive, in modo che le applicazioni autorizzate possano accedere ai dati per conto degli utenti.
2. Applicazione centralizzata: le operazioni di interrogazione delle API di Lake Formation recuperano i dati da Amazon S3 e filtrano i risultati in base a autorizzazioni effettive. Il motore o l'applicazione che si integra con l'operazione dell'API di interrogazione può dipendere da Lake Formation per valutare le autorizzazioni dell'identità chiamante e filtrare in modo sicuro i dati in base a tali autorizzazioni. I motori di query di terze parti vedono e funzionano solo su dati filtrati.

Argomenti

Utilizzo dell'integrazione delle applicazioni Lake Formation

Lake Formation consente ai servizi di terze parti di integrarsi con Lake Formation e ottenere l'accesso temporaneo ai dati di Amazon S3 per conto dei propri utenti tramite l'utilizzo GetTemporaryGlueTableCredentialse GetTemporaryGluePartitionCredentialsle operazioni. Ciò consente ai servizi di terze parti di utilizzare la stessa funzionalità di vendita di autorizzazioni e credenziali utilizzata dagli altri servizi di AWS analisi. Questa sezione descrive come utilizzare queste operazioni API per integrare un motore di query di terze parti con. Lake Formation

Queste operazioni API sono disabilitate per impostazione predefinita. Esistono due opzioni per autorizzare Lake Formation a integrare le applicazioni:

- Configura i tag di sessione IAM che vengono convalidati ogni volta che vengono richiamate le operazioni dell'API di integrazione delle applicazioni

Per ulteriori informazioni, consulta Abilitazione delle autorizzazioni per un motore di query di terze parti per chiamare le operazioni dell'API di integrazione delle applicazioni.

- Abilita l'opzione che consente ai motori esterni di accedere ai dati nelle ubicazioni Amazon S3 con accesso completo alla tabella

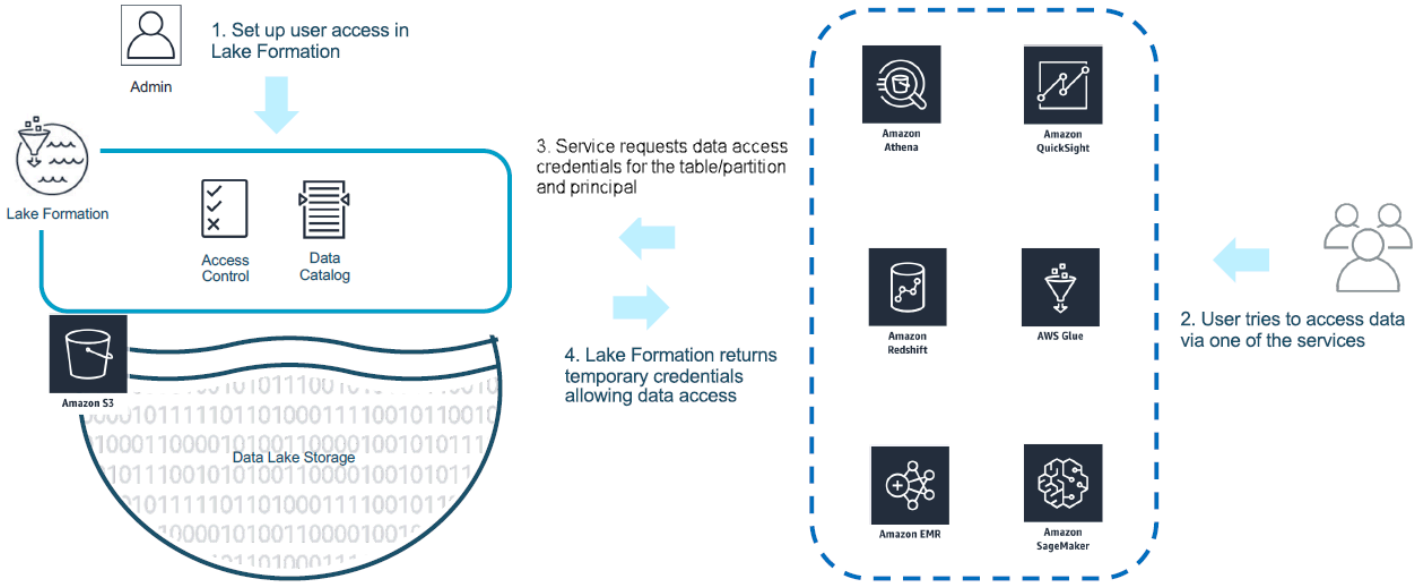
  Questa opzione consente ai motori di query e alle applicazioni di ottenere credenziali senza tag di sessione IAM se l'utente ha accesso completo alla tabella. Fornisce vantaggi in termini di prestazioni ai motori di query e alle applicazioni e semplifica l'accesso ai dati. Amazon EMR su Amazon EC2 è in grado di sfruttare questa impostazione.

  Per ulteriori informazioni, consulta Integrazione delle applicazioni per l'accesso completo alla tabella .

Argomenti

Come funziona l'integrazione delle applicazioni Lake Formation

Questa sezione descrive come utilizzare le operazioni API di integrazione delle applicazioni per integrare un'applicazione di terze parti (motore di query) conLake Formation.



1. L'Lake Formationamministratore svolge le seguenti attività:
   - Registra una sede Amazon S3 con Lake Formation fornendo un ruolo IAM (utilizzato per le credenziali di vendita) con le autorizzazioni appropriate per accedere ai dati all'interno della posizione Amazon S3
   - Registra un'applicazione di terze parti per poter chiamare le operazioni API di vendita delle credenziali di Lake Formation. Per informazioni, consultare the section called “Registrazione di un motore di query di terze parti”.
   - Concede agli utenti le autorizzazioni per accedere a database e tabelle

     Ad esempio, se si desidera pubblicare un set di dati sulle sessioni utente che include alcune colonne contenenti informazioni di identificazione personale (PII), per limitare l'accesso, si assegna a queste colonne un tag LF-TBAC denominato «classificazione» con il valore «sensibile». Successivamente, si definisce un'autorizzazione che consenta a un analista aziendale di accedere ai dati delle sessioni utente, ma si escludono le colonne contrassegnate con classificazione = sensibile.
2. Un principale (utente) invia una richiesta a un servizio integrato.
3. L'applicazione integrata invia la richiesta a Lake Formation chiedendo le informazioni sulla tabella e le credenziali per accedere alla tabella.
4. Se il principale richiedente è autorizzato ad accedere alla tabella, Lake Formation restituisce le credenziali all'applicazione integrata, che consente l'accesso ai dati.

> **Note**
>
> Lake Formation non accede ai dati sottostanti quando vende le credenziali.

5. Il servizio integrato legge i dati da Amazon S3, filtra le colonne in base alle politiche ricevute e restituisce i risultati al principale.

> **Important**
>
> Lake FormationLe operazioni API di vendita delle credenziali abilitano un modello di applicazione distribuita con un modello esplicito di negazione dell'errore (fail-close). Questo introduce un modello di sicurezza a tre parti tra clienti, servizi di terze parti e Lake Formation. I servizi integrati sono affidabili per applicare correttamente le Lake Formation autorizzazioni (applicazione distribuita).

Il servizio integrato è responsabile del filtraggio dei dati letti da Amazon S3 in base alle politiche restituite Lake Formation prima che i dati filtrati vengano restituiti all'utente. I servizi integrati seguono un modello di chiusura fallimentare, il che significa che devono fallire la query se non sono in grado di applicare le autorizzazioni richieste. Lake Formation

Ruoli e responsabilità nell'integrazione delle applicazioni Lake Formation

Ruolo	Responsabilità
Il cliente	• Abilita l'impostazione di integrazione dell'applicazione Lake Formation (vedithe section called “Registrazione di un motore di query di terze parti”). • Registra esplicitamente le terze parti approvate con Lake Formation (vedithe section called “Registrazione di un motore di query di terze parti”). • Testa e convalida soluzioni di terze parti con autorizzazioni Lake Formation. • Monitora e verifica l'utilizzo da parte di terzi delle operazioni dell'API di vendita delle credenziali di Lake Formation.

Ruolo	Responsabilità
La terza parte	• Documenta pubblicamente la funzionalità supportata per ogni revisione del software e fornisce istruzioni per attivarla correttamente. • Pubblicizza in modo accurato le funzionalità supportate quando richiama le operazioni API di vendita delle credenziali di Lake Formation (secondo la documentazione). • Archivia e gestisce in modo sicuro le credenziali vendute per evitare fughe di credenziali e aumento dei privilegi. • Applica le autorizzazioni in base alle funzionalità supportate e restituis ce agli utenti solo dati filtrati • Non riesce a eseguire la query quando non è possibile applicare correttamente le autorizzazioni richieste
`AWS Lake Formation`	• Deriva e restituisce correttamente le autorizzazioni effettive per un determinato principale. • Convalida le funzionalità supportate da terze parti in base al funzionam ento dell'API. call-by-call • Restituisce credenziali IAM limitate solo quando le funzionalità pubblicizzate del motore corrispondono a quelle definite nelle risorse del catalogo, altrimenti restituisce un errore.

Lake Formationflusso di lavoro per le operazioni API di integrazione delle applicazioni

Di seguito è riportato il flusso di lavoro per le operazioni dell'API di integrazione delle applicazioni:

1. Un utente invia una query o una richiesta di dati utilizzando un motore di query integrato di terze parti. Il motore di query assume un ruolo IAM che rappresenta l'utente o un gruppo di utenti e recupera credenziali affidabili da utilizzare quando si chiamano le operazioni dell'API di integrazione delle applicazioni.
2. Il motore di query chiama e`GetUnfilteredTableMetadata`, se si tratta di una tabella partizionata, chiama `GetUnfilteredPartitionsMetadata` per recuperare i metadati e le informazioni sulle policy dal Data Catalog.

3. Lake Formation esegue l'autorizzazione per la richiesta. Se l'utente non dispone delle autorizzazioni appropriate sulla tabella, AccessDeniedExceptionviene generato.
4. Come parte della richiesta, il motore di query invia il filtro che supporta. È possibile inviare due flag all'interno di un array: COLUMN_PERMISSIONS e CELL_FILTER_PERMISSION. Se il motore di query non supporta nessuna di queste funzionalità e nella tabella esiste una policy per la funzionalità, viene generato un messaggio e la query ha esito negativo. PermissionTypeMismatchException Questo serve per evitare la perdita di dati.
5. La risposta restituita contiene quanto segue:
   - L'intero schema della tabella in modo che i motori di query possano utilizzarlo per analizzare i dati dall'archiviazione.
   - Un elenco di colonne autorizzate a cui l'utente ha accesso. Se l'elenco delle colonne autorizzate è vuoto, indica che l'utente dispone `DESCRIBE` delle autorizzazioni ma non `SELECT` delle ha e la query ha esito negativo.
   - Una bandiera`IsRegisteredWithLakeFormation`, che indica se Lake Formation può fornire credenziali ai dati di queste risorse. Se restituisce false, le credenziali dei clienti devono essere utilizzate per accedere ad Amazon S3.
   - Un elenco degli `CellFilters` eventuali elementi da applicare alle righe di dati. Questo elenco contiene colonne e un'espressione per valutare ogni riga. Questo campo deve essere compilato solo se CELL_FILTER_PERMISSION viene inviato come parte della richiesta ed è presente un filtro di dati sulla tabella per l'utente chiamante.
6. Dopo aver recuperato i metadati, il motore di query chiama `GetTemporaryGlueTableCredentials` o `GetTemporaryGluePartitionCredentials` per ottenere AWS le credenziali per recuperare i dati dalla posizione Amazon S3.
7. Il motore di query legge gli oggetti pertinenti da Amazon S3, filtra i dati in base alle politiche ricevute nella fase 2 e restituisce i risultati all'utente.

Le operazioni dell'API di integrazione delle applicazioni Lake Formation contengono contenuti aggiuntivi per la configurazione dell'integrazione con motori di query di terze parti. Puoi vedere i dettagli dell'operazione nella sezione Operazioni dell'API di vendita delle credenziali.

Registrazione di un motore di query di terze parti

Prima che un motore di query di terze parti possa utilizzare le operazioni dell'API di integrazione delle applicazioni, devi abilitare esplicitamente le autorizzazioni affinché il motore di query chiami le operazioni API per tuo conto. Questa operazione viene eseguita in pochi passaggi:

1. È necessario specificare AWS gli account e i tag di sessione IAM che richiedono l'autorizzazione a chiamare le operazioni dell'API di integrazione delle applicazioni tramite la AWS Lake Formation console AWS CLI o l'API/SDK.
2. Quando il motore di query di terze parti assume il ruolo di esecuzione nel tuo account, il motore di query deve allegare un tag di sessione registrato presso Lake Formation che rappresenta il motore di terze parti. Lake Formationutilizza questo tag per verificare che la richiesta provenga da un motore approvato. Per ulteriori informazioni sui tag di sessione, consulta Tag di sessione nella Guida per l'utente IAM.
3. Quando si configura un ruolo di esecuzione di un motore di query di terze parti, è necessario disporre del seguente set minimo di autorizzazioni nella policy IAM:

```
{
  "Version": "2012-10-17",
  "Statement": {"Effect": "Allow",
    "Action": [
      "lakeformation:GetDataAccess",
      "glue:GetTable",
      "glue:GetTables",
      "glue:GetDatabase",
      "glue:GetDatabases",
      "glue:CreateDatabase",
      "glue:GetUserDefinedFunction",
      "glue:GetUserDefinedFunctions",
      "glue:GetPartition",
      "glue:GetPartitions"
    ],
    "Resource": "*"
  }
}
```

4. Imposta una policy di affidabilità dei ruoli sul ruolo di esecuzione del motore di query per avere un controllo di accesso preciso su quale coppia chiave-valore del tag di sessione può essere associata a questo ruolo. Nell'esempio seguente, a questo ruolo è consentito associare solo la chiave del tag di sessione `"LakeFormationAuthorizedCaller"` e il valore `"engine1"` del tag di sessione e non è consentita nessun'altra coppia di valori chiave del tag di sessione.

```
{
    "Sid": "AllowPassSessionTags",
    "Effect": "Allow",
    "Principal": {
```

```
        "AWS": "arn:aws:iam::111122223333:role/query-execution-role"
    },
    "Action": "sts:TagSession",
    "Condition": {
    "StringLike": {
        "aws:RequestTag/LakeFormationAuthorizedCaller": "engine1"        }
    }
}
```

Quando si `LakeFormationAuthorizedCaller` chiama l'operazione STS: AssumeRole API per recuperare le credenziali da utilizzare dal motore di query, il tag di sessione deve essere incluso nella AssumeRole richiesta. La credenziale temporanea restituita può essere utilizzata per effettuare richieste API di integrazione Lake Formation delle applicazioni.

Lake FormationLe operazioni dell'API di integrazione delle applicazioni richiedono che il principale chiamante sia un ruolo IAM. Il ruolo IAM deve includere un tag di sessione con un valore predeterminato che è stato registrato conLake Formation. Questo tag consente di Lake Formation verificare che il ruolo utilizzato per chiamare le operazioni dell'API di integrazione delle applicazioni sia autorizzato a farlo.

Abilitazione delle autorizzazioni per un motore di query di terze parti per chiamare le operazioni dell'API di integrazione delle applicazioni

Segui questi passaggi per consentire a un motore di query di terze parti di richiamare le operazioni dell'API di integrazione delle applicazioni tramite la AWS Lake Formation console, l'AWS CLIAPI/SDK.

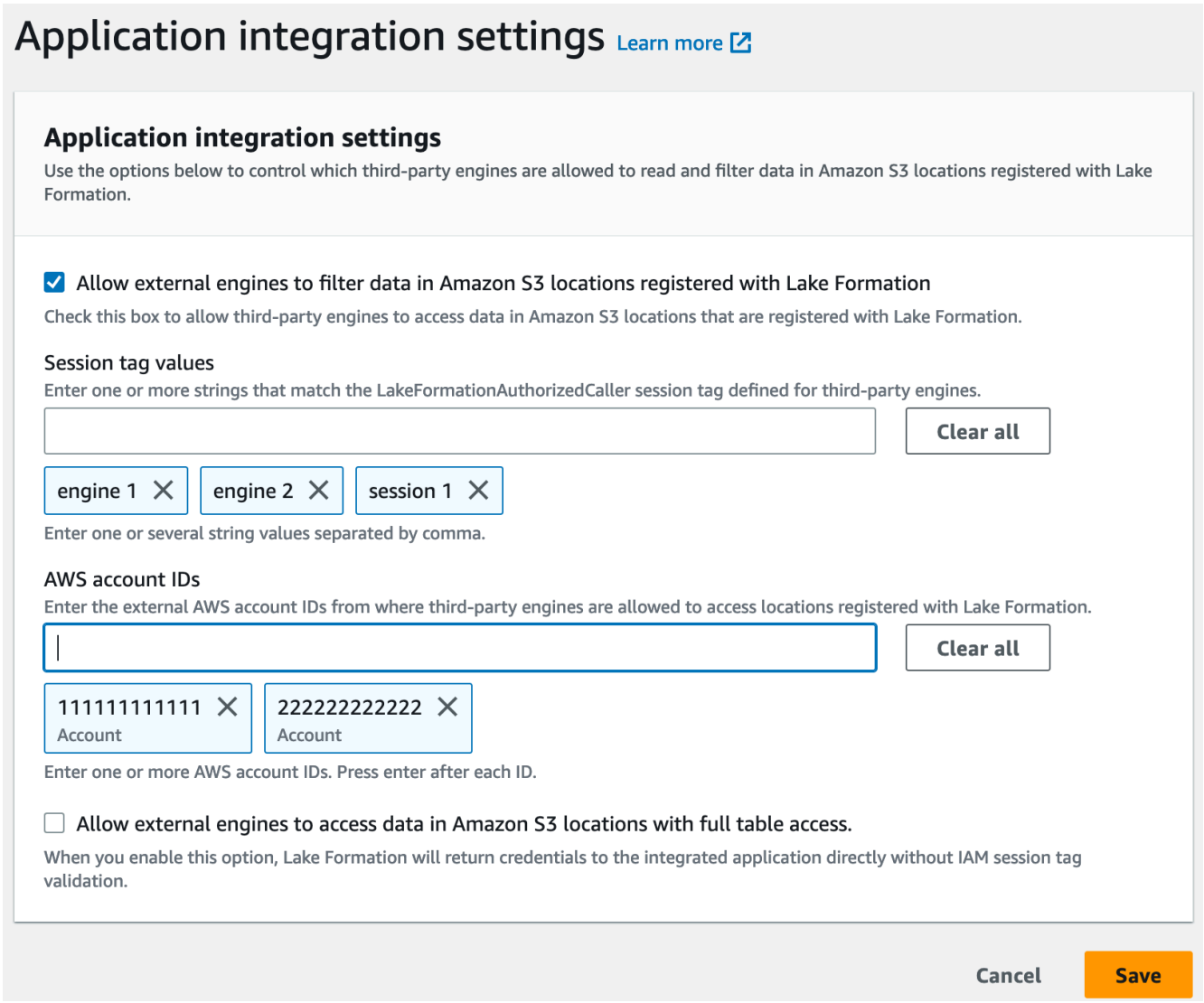
Console

Per registrare il tuo account per il filtraggio dei dati esterni:

1. Accedi a e apri AWS Management Console la console Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.
2. Nella barra di navigazione a sinistra, espandi Autorizzazioni, quindi scegli l'impostazione di integrazione dell'applicazione.
3. Nella pagina delle impostazioni di integrazione delle applicazioni, scegli l'opzione Consenti ai motori esterni di filtrare i dati nelle località Amazon S3 registrate con. Lake Formation

4. Inserisci i tag di sessione che hai creato per il motore di terze parti. Per informazioni sui tag di sessione, consultate Passare i tag di sessione in AWS STS nella Guida AWS Identity and Access Management per l'utente.
5. Immettete gli ID degli account per gli utenti che possono utilizzare il motore di terze parti per accedere a informazioni sui metadati non filtrate e alle credenziali di accesso ai dati delle risorse dell'account corrente.

   Puoi anche utilizzare il campo ID AWS account per configurare l'accesso tra account.



CLI

Utilizzate il comando `put-data-lake-settings` CLI per impostare i seguenti parametri.

Ci sono tre campi da configurare quando si utilizza questo AWS CLI comando:

- `allow-external-data-filtering` — (booleano) Indica che un motore di terze parti può accedere a informazioni sui metadati non filtrate e alle credenziali di accesso ai dati delle risorse dell'account corrente.
- `external-data-filtering-allow-list`— (array) Un elenco di ID di account che possono accedere a informazioni sui metadati non filtrate e alle credenziali di accesso ai dati delle risorse dell'account corrente quando si utilizza un motore di terze parti.
- `authorized-sessions-tag-value-list`— (array) Un elenco di valori (stringhe) dei tag di sessione autorizzati. Se una credenziale di ruolo IAM è stata associata a una coppia chiave-valore autorizzata, se il tag di sessione è incluso nell'elenco, alla sessione viene concesso l'accesso alle informazioni sui metadati non filtrate e alle credenziali di accesso ai dati sulle risorse dell'account configurato. La chiave del tag di sessione autorizzata è definita come. `*LakeFormationAuthorizedCaller*`
- `AllowFullTableExternalDataAccess`- (boolean) Indica se consentire a un motore di query di terze parti di ottenere le credenziali di accesso ai dati senza tag di sessione quando un chiamante dispone delle autorizzazioni complete di accesso ai dati.

Per esempio:

```
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

Utilizzate l'operazione `PutDataLakeSetting` API per impostare i seguenti parametri.

Esistono tre campi da configurare quando si utilizza questa operazione API:

- `AllowExternalDataFiltering`— (Boolean) Indica se un motore di terze parti può accedere a informazioni sui metadati non filtrate e alle credenziali di accesso ai dati delle risorse dell'account corrente.
- `ExternalDataFilteringAllowList`— (array) Un elenco di ID di account che possono accedere a informazioni sui metadati non filtrate e alle credenziali di accesso ai dati delle risorse dell'account corrente utilizzando un motore di terze parti.
- `AuthorizedSectionsTagValueList`— (array) Un elenco di valori di tag autorizzati (stringhe). Se una credenziale di ruolo IAM è stata associata a un tag autorizzato, alla sessione viene concesso l'accesso alle informazioni sui metadati non filtrate e alle credenziali di accesso ai dati sulle risorse dell'account configurato. La chiave del tag di sessione autorizzata è definita come. `*LakeFormationAuthorizedCaller*`
- `AllowFullTableExternalDataAccess`- (boolean) Indica se consentire a un motore di query di terze parti di ottenere le credenziali di accesso ai dati senza tag di sessione quando un chiamante dispone delle autorizzazioni complete di accesso ai dati.

Per esempio:

```
//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));
}
```

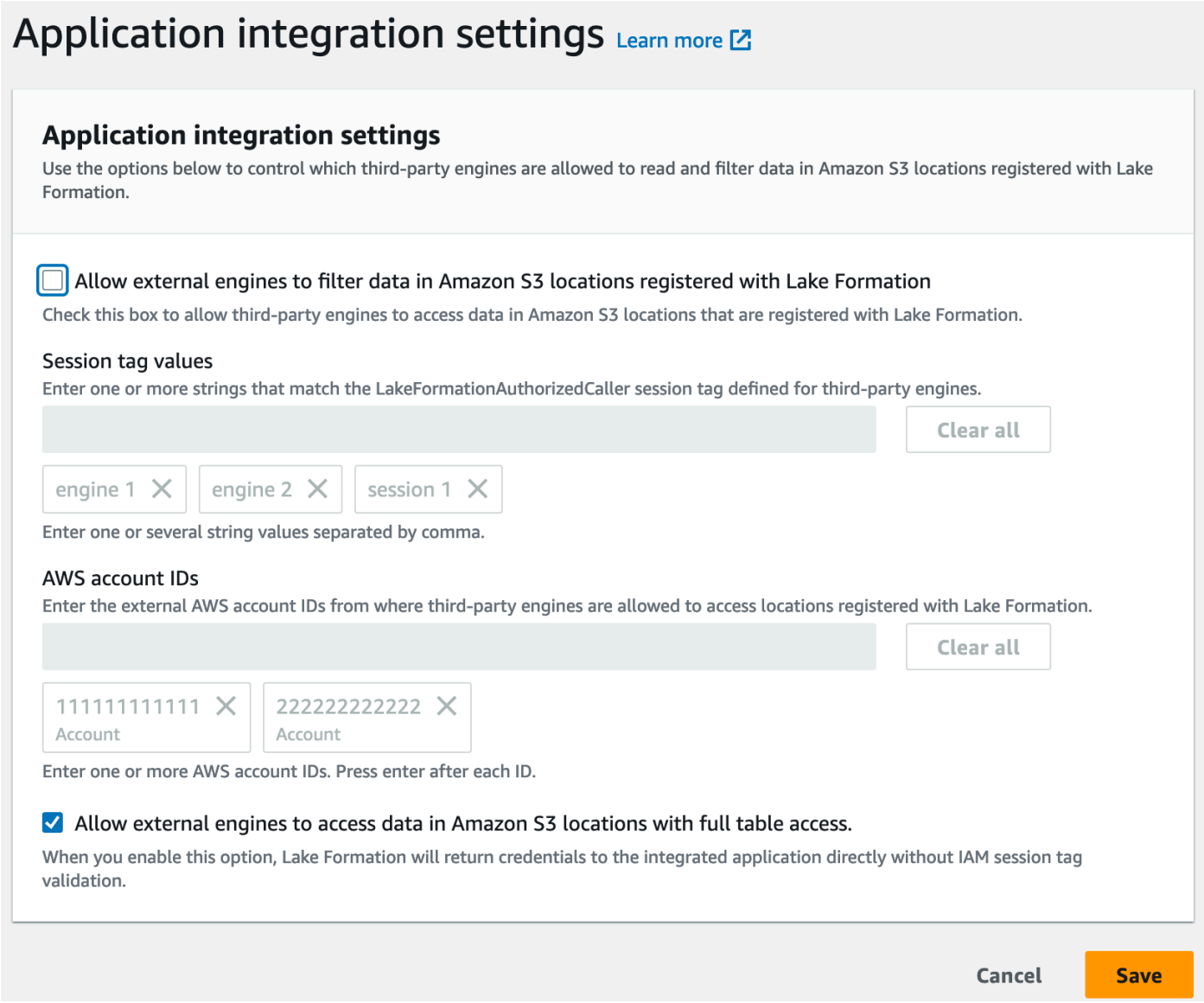
Integrazione delle applicazioni per l'accesso completo alla tabella

Segui questi passaggi per consentire ai motori di query di terze parti di accedere ai dati senza la convalida dei tag di sessione IAM:

Console

1. Accedi alla console di Lake Formation all'indirizzo https://console.aws.amazon.com/lakeformation/.
2. Nella barra di navigazione a sinistra, espandi Autorizzazioni e scegli Impostazioni di integrazione delle applicazioni.
3. Nella pagina delle impostazioni di integrazione dell'applicazione, seleziona la casella di controllo Consenti ai motori esterni di accedere ai dati nelle posizioni Amazon S3 con accesso completo alla tabella.

   Quando abiliti questa opzione, Lake Formation restituirà le credenziali all'applicazione di interrogazione direttamente senza la convalida del tag di sessione IAM.



AWS CLI

Utilizzate il comando `put-data-lake-settings` CLI per impostare il `AllowFullTableExternalDataAccess` parametro.

```
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
    }
}
```

Collaborazione con altri AWS servizi

AWS servizi come Amazon Athena AWS Glue, Amazon Redshift Spectrum e Amazon EMR possono utilizzare Lake Formation per accedere in modo sicuro ai dati nelle sedi Amazon S3 registrate con Lake Formation. Con Lake Formation, puoi definire e gestire le autorizzazioni di controllo degli accessi granulari (FGAC) per i tuoi dati in. AWS Glue Data Catalog Ciascuno di questi AWS servizi è un chiamante affidabile di Lake Formation e Lake Formation fornisce l'accesso ai dati archiviati in Amazon S3 tramite credenziali temporanee. Per ulteriori informazioni, consulta Come funziona l'integrazione delle applicazioni Lake Formation.

Per usufruire di queste funzionalità, Lake Formation richiede innanzitutto la registrazione della posizione Amazon S3 e l'assegnazione delle autorizzazioni appropriate al principale IAM per accedere alla tabella, al database e alla posizione Amazon S3. Per ulteriori informazioni, consulta Gestione delle autorizzazioni di Lake Formation.

Argomenti

Utilizzo AWS Lake Formation con Amazon Athena

Amazon Athena è un servizio di query senza server che ti aiuta ad analizzare dati strutturati, semistrutturati e non strutturati archiviati in Amazon S3. Athena supporta l'interrogazione di dati dai formati di dati CSV, JSON, Parquet e Avro. Athena supporta anche formati di tabella come le tabelle governate da Apache Hive, Apache Hudi, Apache Iceberg e Lake Formation. Athena si integra con l' AWS Glue Data Catalog archiviazione dei metadati dei tuoi set di dati in Amazon S3. Athena può utilizzare Lake Formation per definire e mantenere le politiche di controllo degli accessi su tali set di dati.

Ecco alcuni casi d'uso comuni in cui è possibile utilizzare Lake Formation con Athena.

- Utilizza le autorizzazioni di Lake Formation per accedere alle risorse del Data Catalog (database e tabelle) da Athena. È possibile utilizzare il metodo della risorsa denominata o LF-Tags per definire le autorizzazioni su database e tabelle. Per ulteriori informazioni, consultare:
  - Concessione delle autorizzazioni al database utilizzando il metodo di risorsa denominato
  - Controllo degli accessi basato su tag Lake Formation

  > **Note**
  >
  > Le autorizzazioni di Lake Formation si applicano solo quando si utilizza Athena per interrogare i dati di origine da Amazon S3 e i metadati in Data Catalog.

  Le autorizzazioni di Lake Formation supportano operazioni di lettura e scrittura su database e tabelle.

  > **Note**
  >
  > Non puoi applicare filtri di dati quando usi LF-Tags per gestire le autorizzazioni sulle risorse di Data Catalog.

- Controlla i risultati delle query utilizzando Filtri di dati in Lake Formation per proteggere le tabelle nei tuoi data lake Amazon S3 concedendo autorizzazioni a livello di colonna, riga e cella. Consulta la limitazione alla proiezione delle partizioni nella Guida per l'utente di Amazon Athena.
- Applica un controllo granulare degli accessi sui dati disponibili per l'utente Athena basato su SAML durante l'esecuzione di query federate.

  I driver JDBC e ODBC Athena supportano la configurazione dell'accesso federato all'origine dati utilizzando Identity Provider (IdP) basato su SAML. Usa Amazon QuickSight integrato con Lake Formation con il tuo ruolo IAM esistente o con utenti o gruppi SAML per visualizzare i risultati delle query Athena.

  > **Note**
  >
  > Le autorizzazioni di Lake Formation per utenti e gruppi SAML verranno applicate solo quando invii query ad Athena utilizzando il driver JDBC o ODBC.

Per ulteriori informazioni, consulta Utilizzo dei driver Lake Formation e Athena JDBC e ODBC per l'accesso federato ad Athena.

> **Note**
>
> Attualmente, l'autorizzazione dell'accesso alle identità SAML in Lake Formation non è supportata nelle seguenti regioni:
>
> - Medio Oriente (Bahrein) - me-south-1
> - Asia Pacifico (Hong Kong) - ap-east-1
> - Africa (Città del Capo) - af-south-1
> - Cina (Ningxia) - cn-nordovest-1
> - Asia Pacifico (Osaka) - ap-northeast-3

- Condivisione dei dati tra account in Lake FormationDa utilizzare per interrogare le tabelle in un altro account.

> **Note**
>
> Per ulteriori informazioni sulle limitazioni relative all'utilizzo delle autorizzazioni di Lake Formation per`Views`, consulta Considerazioni e limitazioni.

Support per i formati di tabelle transazionali

L'applicazione delle autorizzazioni Lake Formation consente di proteggere i dati transazionali nei data lake basati su Amazon S3. La tabella seguente elenca i formati di tabelle transazionali supportati nelle autorizzazioni Athena e Lake Formation. Lake Formation applica queste autorizzazioni quando gli utenti di Athena eseguono le loro query.

Formato della tabella	Descrizione e operazioni consentite	Autorizzazioni Lake Formation supportate in Athena
Apache Hudi	Un formato utilizzato per semplificare l'elaborazione	Utilizzalo Filtraggio dei dati e sicurezza a livello di cella in Lake Formation per protegger

Formato della tabella	Descrizione e operazioni consentite	Autorizzazioni Lake Formation supportate in Athena
	incrementale dei dati e lo sviluppo di pipeline di dati. Athena supporta operazioni di creazione e lettura utilizzando i formati di tabella Apache Hudi su set di dati Amazon S3 per i tipi di tabella Hudi Copy on Write (CoW) e Merge On Read (MoR). Athena non supporta le operazioni di scrittura sulle tabelle Hudi. Usa Athena per interrogare i set di dati Hudi.	e la tabella Hudi utilizzando le autorizzazioni a livello di tabella, colonna, riga e cella.
Apache Iceberg	Un formato di tabella aperta che gestisce grandi raccolte di file sotto forma di tabelle e supporta le moderne operazioni analitiche dei data lake come l'inserimento, l'aggiornamento, l'eliminazione e le query sui viaggi nel tempo a livello di record. Per ulteriori informazioni sul supporto di Athena per le tabelle Iceberg, vedere Utilizzo delle tabelle Iceberg.	Sono supportate le autorizzazioni a livello di tabella, colonna, riga e cella. Attualmente, Lake Formation non supporta la gestione delle autorizzazioni su operazioni di scrittura come `VACUUM` `UPDATE` e `OPTIMIZE` sulle tabelle in Open Table Formats. `MERGE`

Formato della tabella	Descrizione e operazioni consentite	Autorizzazioni Lake Formation supportate in Athena
Linux Foundation Delta Lake	Delta Lake è un progetto open source che aiuta a implement are moderne architetture di data lake comunemen te costruite su Amazon S3 o Hadoop Distributed File System (HDFS). Athena supporta le tabelle Delta Lake create utilizzan do una definizione di tabella manifest basata su symlink a AWS Glue Data Catalog partire da una tabella Delta Lake. Per ulteriori informazioni, consulta Scansionare le tabelle Delta Lake utilizzando i crawler. AWS Glue Athena (versione 3 del motore) supporta la lettura delle tabelle native di Delta Lake. Per ulteriori informazioni, consulta Introduzione al supporto tabellare nativo di Delta Lake con i AWS Glue crawler.	Le autorizzazioni a livello di tabella, colonna, riga e cella sono supportate per le tabelle symlink e le tabelle native di Delta Lake.

Risorse aggiuntive

Post di blog, video e workshop

- Interroga un set di dati Apache Hudi in un data lake Amazon S3 con Amazon Athena
- Crea un data lake Apache Iceberg utilizzando Amazon Athena, Amazon EMR e AWS Glue
- Inserisci, aggiorna, elimina su Amazon S3 con Athena e Apache Iceberg
- Workshop Lake Formation sul controllo degli accessi basato su LF-tag sull'interrogazione di un data lake.

Utilizzo AWS Lake Formation con Amazon Redshift Spectrum

Amazon Redshift Spectrum ti consente di interrogare e recuperare dati nei data lake Amazon S3 senza caricare dati nei nodi del cluster Amazon Redshift.

Redshift Spectrum supporta due modi per registrare un catalogo di AWS Glue dati esterno abilitato con Lake Formation.

- Utilizzo di un ruolo IAM collegato al cluster con autorizzazione al Data Catalog

  Per creare un ruolo IAM, segui i passaggi descritti nella procedura seguente.

  Per creare un ruolo IAM per Amazon Redshift utilizzando un AWS Glue Data Catalog enabled for AWS Lake Formation
- Utilizzo di un'identità IAM federata configurata per gestire l'accesso a risorse esterne AWS Glue Data Catalog

  Redshift Spectrum supporta l'interrogazione delle tabelle di Lake Formation utilizzando identità IAM federate. Le identità IAM possono essere un utente IAM o un ruolo IAM. Per ulteriori informazioni sulla federazione delle identità IAM in Redshift Spectrum, consulta Usare un'identità federata per gestire l'accesso di Amazon Redshift alle risorse locali e le tabelle esterne di Redshift Spectrum.

Grazie all'integrazione di Lake Formation con Redshift Spectrum, puoi definire le autorizzazioni di controllo degli accessi a livello di riga, colonna e cella sulle tabelle dopo la registrazione dei dati con Lake Formation.

Per ulteriori informazioni, consulta Usare Redshift Spectrum con. AWS Lake Formation

Redshift Spectrum supporta letture o `SELECT` interrogazioni sulle tabelle di schemi esterne gestite da Lake Formation.

Per ulteriori informazioni, consulta Creazione di schemi esterni per Redshift Spectrum.

Support per tipi di tabelle transazionali

Questa tabella elenca i formati di tabelle transazionali supportati in Redshift Spectrum e le autorizzazioni Lake Formation applicabili.

Formati di tabella supportati

Formato della tabella	Descrizione e operazioni consentite	Autorizzazioni Lake Formation supportate in Redshift Spectrum
Apache Hudi	Un formato utilizzato per semplificare l'elaborazione incrementale dei dati e lo sviluppo di pipeline di dati. Redshift Spectrum supporta le operazioni di inserimento, eliminazione e storno di scrittura utilizzando il formato di tabella Apache Hudi Copy on Write (CoW) su Amazon S3. Per ulteriori informazioni, consulta Creazione di tabelle esterne per i dati gestiti in Apache Hudi.	Filtraggio dei dati e sicurezza a livello di cella in Lake FormationUtilizzatelo per proteggere le tabelle Hudi utilizzando le autorizzazioni a livello di tabella, colonna, riga e cella.
Apache Iceberg	Un formato di tabella aperta che gestisce grandi raccolte di file sotto forma di tabelle e supporta le moderne operazioni analitiche di data	Redshift Spectrum supporta le tabelle Apache Iceberg per le interrogazioni.

Formato della tabella	Descrizione e operazioni consentite	Autorizzazioni Lake Formation supportate in Redshift Spectrum
	lake come l'inserimento, l'aggiornamento, l'eliminazione e le query sui viaggi nel tempo a livello di record. Per ulteriori informazioni, consulta Usare le tabelle Apache Iceberg con Amazon Redshift.	
Linux Foundation Delta Lake	Delta Lake è un progetto open source che aiuta a implement are moderne architetture di data lake comunemen te costruite su Amazon S3 o Hadoop Distributed File System (HDFS). Redshift Spectrum supporta l'interrogazione delle tabelle Delta Lake. Per ulteriori informazioni, consulta Creazione di tabelle esterne per i dati gestiti in Delta Lake.	Sono supportate le autorizza zioni a livello di tabella, colonna, riga e cella.

Risorse aggiuntive

Post di blog e workshop

- Centralizza la governance per il tuo data lake utilizzando, al AWS Lake Formation contempo, un'architettura di dati moderna con Amazon Redshift Spectrum
- Usa Redshift Spectrum per interrogare le tabelle Apache HUDI Copy On Write (CoW) nel data lake Amazon S3

AWS Lake Formation Utilizzo con AWS Glue

I data engineer e i DevOps professionisti utilizzano AWS Glue Extract, Transform and Load (ETL) con Apache Spark per eseguire trasformazioni sui propri set di dati in Amazon S3 e caricare i dati trasformati in data lake e data warehouse per analisi, apprendimento automatico e sviluppo di applicazioni. Poiché diversi team accedono allo stesso set di dati in Amazon S3, è fondamentale concedere e limitare le autorizzazioni in base ai rispettivi ruoli.

AWS Lake Formation è basato su e AWS Glue i servizi interagiscono nei seguenti modi:

- Lake Formation e AWS Glue condividono lo stesso Data Catalog.
- Le seguenti funzionalità della console di Lake Formation richiamano la AWS Glue console:
  - Lavori — Per ulteriori informazioni, consulta Aggiungere lavori nella Guida per gli AWS Glue sviluppatori.
  - Crawler — Per ulteriori informazioni, consulta Cataloging Tables with a Crawler nella Developer Guide.AWS Glue
- I flussi di lavoro generati quando si utilizza un blueprint di Lake Formation sono AWS Glue flussi di lavoro. Puoi visualizzare e gestire questi flussi di lavoro sia nella console di Lake Formation che nella AWS Glue console.
- Le trasformazioni di machine learning sono fornite con Lake Formation e si basano su operazioni AWS Glue API. Puoi creare e gestire le trasformazioni dell'apprendimento automatico sulla AWS Glue console. Per ulteriori informazioni, consulta Machine Learning Transforms nella AWS Glue Developer Guide.

Puoi utilizzare il controllo granulare degli accessi di Lake Formation per gestire le risorse del Data Catalog esistenti e le posizioni dati Amazon S3.

> **Note**
>
> AWS Glue ETL richiede l'accesso completo all'intera tabella durante il recupero dei dati dalla posizione Amazon S3 sottostante. AWS Glue Il processo ETL ha esito negativo se si applicano autorizzazioni a livello di colonna su una tabella. Tuttavia, è possibile creare una sicurezza a livello di colonna e di riga definendo filtri di dati. Per ulteriori informazioni, consulta Note e restrizioni per il filtraggio a livello di colonna Lake Formation valuta il filtro dei dati definito nella tabella e recupera solo i dati filtrati da Amazon S3 necessari per il job ETL. AWS Glue

Support per tipi di tabelle transazionali

L'applicazione delle autorizzazioni Lake Formation consente di proteggere i dati transazionali nei data lake basati su Amazon S3. La tabella seguente elenca i formati di tabelle transazionali supportati AWS Glue e le autorizzazioni di Lake Formation. Lake Formation applica queste autorizzazioni per AWS Glue le operazioni.

Formati di tabella supportati

Formato della tabella	Descrizione e operazioni consentite	Autorizzazioni Lake Formation supportate in AWS Glue
Apache Hudi	Un formato di tabella aperta utilizzato per semplificare l'elaborazione increment ale dei dati e lo sviluppo di pipeline di dati. Per esempi, vedete Using the Hudi framework in. AWS Glue	Le autorizzazioni a livello di tabella sono disponibili per le tabelle Hudi. Per ulteriori informazioni, consulta Limitazioni.
Apache Iceberg	Un formato di tabella aperta che gestisce grandi raccolte di file come tabelle. Per esempi, vedete Using the Iceberg framework in AWS Glue.	Le autorizzazioni a livello di tabella sono disponibili per le tabelle Iceberg. Per ulteriori informazioni, consulta Limitazioni.
Linux Foundation Delta Lake	Delta Lake è un progetto open source che aiuta a implement are moderne architetture di data lake comunemen te costruite su Amazon S3 o Hadoop Distributed File System (HDFS).	Le autorizzazioni a livello di tabella sono disponibili per le tabelle Delta Lake. Per ulteriori informazioni, consulta Limitazioni.

Formato della tabella	Descrizione e operazioni consentite	Autorizzazioni Lake Formation supportate in AWS Glue
	Per esempi, consulta Usare il framework Delta Lake in. AWS Glue	

Risorse aggiuntive

Post e repository del blog

- Usa il AWS Glue connettore per leggere e scrivere tabelle Apache Iceberg con transazioni ACID ed eseguire viaggi nel tempo
- Scrittura su tabelle Apache Hudi utilizzando un connettore personalizzato AWS Glue
- AWS repository del modello Cloudformation e dell'esempio di codice pyspark per analizzare i dati di streaming utilizzando Apache Hudi e AWS Glue Amazon S3.

Utilizzo AWS Lake Formation con Amazon EMR

Amazon EMR è una piattaforma cluster AWS gestita flessibile su cui è possibile eseguire qualsiasi codice personalizzato su framework di big data supportati come Hadoop Map-Reduce, Spark, Hive, Presto, ecc. Le organizzazioni utilizzano Amazon EMR anche per eseguire applicazioni di elaborazione dati in batch e in streaming su un cluster altamente distribuito. Con Amazon EMR, puoi eseguire trasformazioni di dati e codice personalizzato su database e tabelle le cui autorizzazioni sono gestite da Lake Formation.

Esistono tre opzioni per la distribuzione di Amazon EMR:

- EMR su EC2
- EMR serverless
- Amazon EMR su EKS

Per ulteriori informazioni, consulta Integrazione di Amazon EMR con Lake Formation o Utilizzo di EMR Serverless con per un controllo granulare degli accessi AWS Lake Formation

Support per i formati di tabelle transazionali

Le versioni 6.15.0 e successive di Amazon EMR includono il supporto per le autorizzazioni di controllo degli accessi a livello di tabella, riga, colonna e cella di Lake Formation sui formati di tabella Apache Hudi, Apache Iceberg e Delta Lake quando leggi e scrivi dati con Spark SQL.

Formati di tabella supportati

Formato della tabella	Descrizione e operazioni consentite	Autorizzazioni Lake Formation supportate in Amazon EMR
Apache Hudi	Un formato di tabella aperta utilizzato per semplificare l'elaborazione increment ale dei dati e lo sviluppo di pipeline di dati. Per un elenco delle operazion i supportate, consulta Apache Hudi e Lake Formation.	Amazon EMR supporta il controllo degli accessi a livello di tabella, riga, colonna e cella con Apache Hudi.
Apache Iceberg	Un formato di tabella aperta che gestisce grandi raccolte di file sotto forma di tabelle. Per un elenco delle operazion i supportate, consulta Apache Iceberg e Lake Formation.	Amazon EMR supporta il controllo degli accessi a livello di tabella, riga, colonna e cella con Apache Iceberg.
Linux Foundation Delta Lake	Delta Lake è un progetto open source che aiuta a implement are moderne architetture di data lake comunemen te costruite su Amazon S3 o Hadoop Distributed File System (HDFS).	Amazon EMR supporta il controllo degli accessi a livello di tabella, riga, colonna e cella con le tabelle Delta Lake.

Formato della tabella	Descrizione e operazioni consentite	Autorizzazioni Lake Formation supportate in Amazon EMR
	Per un elenco delle operazioni supportate, consulta Delta Lake and Lake Formation.	

Risorse aggiuntive

Guida per l'utente, post di blog e workshop

- Integrazione con Amazon EMR tramite Runtime Roles
- Inizia rapidamente a usare Apache Hudi, Apache Iceberg e Delta Lake con Amazon EMR su EKS
- Utilizzo di Delta Lake OSS con EMR Serverless

Utilizzo AWS Lake Formation con Amazon QuickSight

Amazon QuickSight supporta l'esplorazione di set di dati gestiti dalle autorizzazioni Lake Formation in Amazon S3 utilizzando Athena.

Sia gli utenti delle edizioni Standard che Enterprise di Amazon si QuickSight integrano con Lake Formation, ma in modo leggermente diverso.

- Enterprise edition: concedi autorizzazioni granulari di controllo degli accessi (FGAC) a singoli QuickSight utenti, gruppi e ruoli IAM di Amazon per accedere a database e tabelle.
- Edizione standard: concedi le autorizzazioni ai ruoli IAM per accedere a database e tabelle.

> **Note**
>
> Per impostazione predefinita, Amazon QuickSight utilizza un ruolo denominato`aws-quicksight-service-role-v0`. Puoi anche definire ruoli personalizzati con le autorizzazioni necessarie che consentono QuickSight ad Amazon di accedere ad Athena.

Per ulteriori informazioni, consulta Autorizzazione delle connessioni tramite AWS Lake Formation

Risorse aggiuntive

Post del blog

- Abilita autorizzazioni dettagliate per gli autori di Amazon in QuickSight AWS Lake Formation
- Analizza in modo sicuro i tuoi dati con AWS Lake Formation Amazon QuickSight

Utilizzo AWS Lake Formation con Lake AWS CloudTrail

AWS CloudTrail Lake supporta l'esplorazione degli archivi di dati di eventi utilizzando Amazon Athena autorizzazioni granulari in. AWS Lake Formation

> **Note**
>
> CloudTrail Lake può essere solo interrogato. Amazon Athena

Per registrare il tuo archivio dati di eventi CloudTrail Lake con Lake Formation, consulta Federare un data store di eventi.

Registrazione di logAWSChiamate API Lake Formation utilizzandoAWS CloudTrail

AWSLake Formation è integrato conAWS CloudTrail, un servizio che fornisce una registrazione delle operazioni eseguite da un utente, un ruolo o unAWSservizio in Lake Formation. CloudTrail acquisisce tutte le chiamate API Lake Formation come eventi. Le chiamate acquisite includono chiamate dalla console Lake Formation, ilAWS Command Line Interfacee chiamate in codice alle azioni dell'API Lake Formation. Se si crea un trail, è possibile abilitare la distribuzione continua di eventi CloudTrail in un bucket Amazon S3, inclusi gli eventi per Lake Formation. Se non si configura un trail, è comunque possibile visualizzare gli eventi più recenti nella console di CloudTrail in Event history (Cronologia eventi). Utilizzando le informazioni raccolte da CloudTrail, è possibile determinare la richiesta effettuata a Lake Formation, l'indirizzo IP da cui la richiesta è stata eseguita, l'autore della richiesta, il momento in cui è stata eseguita e altri dettagli.

Per ulteriori informazioni su CloudTrail, consultare la Guida per l'utente di AWS CloudTrail.

Informazioni sulla Lake Formation in CloudTrail

CloudTrail è abilitato per impostazione predefinita al momento della creazione di un nuovoAWSconto. Quando si verifica un'attività in Lake Formation, questa viene registrata come evento CloudTrail insieme ad altri.AWSeventi dei servizi inCronologia eventi. Un evento rappresenta una singola richiesta da un'origine e include informazioni sull'operazione richiesta, data e ora dell'operazione e parametri della richiesta. Inoltre, ogni evento o voce di log contiene informazioni sull'utente che ha generato la richiesta. Le informazioni di identità consentono di determinare quanto segue:

- Se la richiesta è stata effettuata con credenziali utente root o AWS Identity and Access Management (IAM).
- Se la richiesta è stata effettuata con le credenziali di sicurezza temporanee per un ruolo o un utente federato.
- Se la richiesta è stata effettuata da un altro servizio AWS.

Per ulteriori informazioni, consulta Elemento CloudTrail userIdentity.

Puoi visualizzare, ricercare e scaricare eventi recenti per l'account AWS. Per ulteriori informazioni, consulta Visualizzazione di eventi mediante la cronologia eventi di CloudTrail.

Per una registrazione continuativa di eventi nella tuaAWSaccount, inclusi gli eventi per Lake Formation, crea un trail. Un trail consente a CloudTrail di distribuire i file di log in un bucket Amazon S3. Per impostazione predefinita, quando crei un trail nella console, il trail si applica a tutte le regioni AWS. Il trail registra gli eventi di tutte le regioni nella partizione AWS e distribuisce i file di registro nel bucket Amazon S3 specificato. Inoltre, è possibile configurare altri.AWSservizi, comeAmazon Athenaper analizzare con maggiore dettaglio e utilizzare i dati raccolti nei log CloudTrail. CloudTrail può anche distribuire file di log ad Amazon CloudWatch Logs e CloudWatch Events.

Per ulteriori informazioni, consulta gli argomenti seguenti:

- Panoramica della creazione di un percorso
- CloudTrail supported services and integrations (Servizi e integrazioni CloudTrail supportati)
- Configurazione delle notifiche Amazon SNS per CloudTrail
- Ricezione di file di registro CloudTrail da più Regioni e Ricezione di file di log CloudTrail da più account

Eventi Formation Lake Formation

Tutte le operazioni API Lake Formation vengono registrate da CloudTrail e sono documentate nellaAWS Lake FormationGuida per sviluppatori di . Ad esempio, le chiamate alle operazioni `PutDataLakeSettings`, `GrantPermissions` e `RevokePermissions` generano voci nei file di log di CloudTrail.

L'esempio seguente mostra un evento CloudTrail per`GrantPermissions`Operazione . La voce include l'utente che ha concesso l'autorizzazione (`datalake_admin`), il principale a cui è stata concessa l'autorizzazione (`datalake_user1`) e il permesso concesso (`CREATE_TABLE`). La voce mostra inoltre che la concessione non è riuscita perché il database di destinazione non è stato specificato nel`resource`eargomentazione.

```
{
    "eventVersion": "1.08",
    "userIdentity": {
        "type": "IAMUser",
        "principalId": "AIDAZKE67KM3P775X74U2",
        "arn": "arn:aws:iam::111122223333:user/datalake_admin",
        "accountId": "111122223333",
        "accessKeyId": "...",
        "userName": "datalake_admin"
```

```
    },
    "eventTime": "2021-02-06T00:43:21Z",
    "eventSource": "lakeformation.amazonaws.com",
    "eventName": "GrantPermissions",
    "awsRegion": "us-east-1",
    "sourceIPAddress": "72.21.198.65",
    "userAgent": "aws-cli/1.19.0 Python/3.6.12
 Linux/4.9.230-0.1.ac.223.84.332.metal1.x86_64 botocore/1.20.0",
    "errorCode": "InvalidInputException",
    "errorMessage": "Resource must have one of the have either the catalog, table or
 database field populated.",
    "requestParameters": {
        "principal": {
            "dataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"
        },
        "resource": {},
        "permissions": [
            "CREATE_TABLE"
        ]
    },
    "responseElements": null,
    "requestID": "b85e863f-e75d-4fc0-9ff0-97f943f706e7",
    "eventID": "8d2ccef0-55f3-42d3-9ede-3a6faedaa5c1",
    "readOnly": false,
    "eventType": "AwsApiCall",
    "managementEvent": true,
    "eventCategory": "Management",
    "recipientAccountId": "111122223333"
}
```

L'esempio seguente mostra una voce di log di CloudTrail per`GetDataAccess`Operazione . I principali non chiamano direttamente questa API. Piuttosto,`GetDataAccess`viene registrato ogni volta che un principale o integratoAWSrichiede credenziali temporanee per accedere ai dati in una posizione data lake registrata con 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"
    },
...
}
```

Vedi anche

- Registrazione su più account CloudTrail

Le migliori pratiche, considerazioni e limitazioni di Lake Formation

Usa questa sezione per trovare rapidamente le migliori pratiche, le considerazioni e le limitazioni all'interno. AWS Lake Formation

Consulta le quote di servizio per il numero massimo di risorse o operazioni di servizio per la tua. Account AWS

Argomenti

Buone pratiche e considerazioni sulla condivisione dei dati tra account

Le funzionalità cross-account di Lake Formation consentono agli utenti di condividere in modo sicuro i data lake distribuiti tra più AWS organizzazioni o direttamente con i responsabili IAM in un altro account Account AWS, fornendo un accesso granulare ai metadati del Data Catalog e ai dati sottostanti.

Prendi in considerazione le seguenti best practice per l'utilizzo della condivisione di dati tra account di Lake Formation:

- Non c'è limite al numero di concessioni di autorizzazioni di Lake Formation che puoi concedere ai mandanti per tuo conto. AWS Tuttavia, Lake Formation utilizza la capacità AWS Resource Access

Manager (AWS RAM) per le sovvenzioni tra account che il tuo account può effettuare con il metodo delle risorse denominato. Per massimizzare la AWS RAM capacità, segui queste migliori pratiche per il metodo delle risorse denominate:

- Utilizza la nuova modalità di concessione tra account (versione 3 e successive nelle impostazioni della versione Cross account) per condividere una risorsa con un esterno. Account AWS Per ulteriori informazioni, consulta Aggiornamento delle impostazioni della versione di condivisione dei dati tra account.
- Suddividi AWS gli account in organizzazioni e concedi le autorizzazioni alle organizzazioni o alle unità organizzative. Una concessione a un'organizzazione o a un'unità organizzativa conta come un'unica concessione.

  La concessione a organizzazioni o unità organizzative elimina inoltre la necessità di accettare un AWS Resource Access Manager (AWS RAM) invito alla condivisione di risorse per la sovvenzione. Per ulteriori informazioni, consulta Accesso e visualizzazione di tabelle e database condivisi del Data Catalog.
- Invece di concedere le autorizzazioni su molte singole tabelle di un database, utilizzate la speciale jolly Tutte le tabelle per concedere le autorizzazioni su tutte le tabelle del database. La concessione su tutte le tabelle viene considerata un'unica concessione. Per ulteriori informazioni, consulta Concessione e revoca delle autorizzazioni per le risorse del Data Catalog.

> **Note**
>
> Per ulteriori informazioni sulla richiesta di un limite più elevato per il numero di condivisioni di risorse in AWS RAM, vedere le quote AWS di servizio nel. Riferimenti generali di AWS

- È necessario creare un link di risorsa a un database condiviso affinché tale database appaia negli Amazon Athena editor di query di Amazon Redshift Spectrum. Analogamente, per poter eseguire query su tabelle condivise utilizzando Athena e Redshift Spectrum, è necessario creare collegamenti di risorse alle tabelle. I collegamenti alle risorse vengono quindi visualizzati nell'elenco delle tabelle degli editor di query.

  Invece di creare collegamenti alle risorse per molte singole tabelle per eseguire interrogazioni, è possibile utilizzare il carattere jolly Tutte le tabelle per concedere le autorizzazioni su tutte le tabelle di un database. Quindi, quando crei un link a una risorsa per quel database e lo selezioni nell'editor di query, avrai accesso a tutte le tabelle del database per la tua query. Per ulteriori informazioni, consulta Creazione di collegamenti alle risorse.

- Quando condividi risorse direttamente con i responsabili di un altro account, il responsabile IAM nell'account del destinatario potrebbe non avere l'autorizzazione a creare collegamenti alle risorse per poter interrogare le tabelle condivise utilizzando Athena e Amazon Redshift Spectrum. Invece di creare un link a una risorsa per ogni tabella condivisa, l'amministratore del data lake può creare un database segnaposto e concedere `CREATE_TABLE` l'autorizzazione al gruppo. `ALLIAMPrincipal` Quindi, tutti i responsabili IAM presenti nell'account del destinatario possono creare collegamenti alle risorse nel database dei segnaposto e iniziare a interrogare le tabelle condivise.

  Vedi il comando CLI di esempio per concedere le autorizzazioni a in. `ALLIAMPrincipals` Concessione delle autorizzazioni al database utilizzando il metodo di risorsa denominato
- Athena e Redshift Spectrum supportano il controllo degli accessi a livello di colonna, ma solo per l'inclusione, non per l'esclusione. Il controllo degli accessi a livello di colonna non è supportato nei job ETL. AWS Glue
- Quando una risorsa viene condivisa con il tuo AWS account, puoi concedere le autorizzazioni sulla risorsa solo agli utenti del tuo account. Non puoi concedere le autorizzazioni sulla risorsa ad altri AWS account, alle organizzazioni (nemmeno alla tua organizzazione) o al `IAMAllowedPrincipals` gruppo.
- Non puoi concedere `DROP` o `Super` su un database a un account esterno.
- Revoca le autorizzazioni tra account prima di eliminare un database o una tabella. Altrimenti, è necessario eliminare le condivisioni di risorse orfane in. AWS Resource Access Manager

> Consulta anche
>
> - Buone pratiche e considerazioni per il controllo degli accessi basato su tag Lake Formation
> - `CREATE_TABLE`nella sezione Riferimento alle autorizzazioni di Lake Formation per ulteriori regole e limitazioni di accesso tra account.

Limitazioni di accesso ai dati tra regioni

Lake Formation supporta l'interrogazione su tutte Regioni AWS le tabelle del Data Catalog. Puoi accedere ai dati in una regione da altre regioni utilizzando Amazon Athena Amazon EMR ed AWS Glue ETL creando collegamenti di risorse in altre regioni che puntano ai database e alle tabelle di

origine. Con l'accesso alle tabelle tra regioni, puoi accedere ai dati di tutte le regioni senza copiare i dati o i metadati sottostanti nel Data Catalog.

Le seguenti limitazioni si applicano all'accesso alle tabelle tra aree geografiche.

- Lake Formation non supporta l'interrogazione di tabelle Data Catalog da un'altra regione utilizzando Amazon Redshift Spectrum.
- Nella console Lake Formation, le viste del database e della tabella non mostrano i nomi dei database/tabelle Region di origine.
- Per visualizzare l'elenco delle tabelle in un database condiviso di un'altra regione, devi prima creare un collegamento di risorsa al database condiviso, quindi selezionare il collegamento alla risorsa e scegliere Visualizza tabelle.
- La funzionalità di accesso alle tabelle tra aree geografiche non funziona quando si creano collegamenti a risorse Regioni AWS che puntano a database e tabelle condivisi creati nelle regioni opt-in.

  Per ulteriori informazioni, consulta Opt-in Regions nella pagina Regioni AWS Supportate e servizi.
- Lake Formation non supporta le chiamate ai link di risorse interregionali effettuate dagli utenti SAML.

Data Catalog visualizza, considerazioni e limitazioni

In AWS Glue Data Catalog, una vista è una tabella virtuale in cui i contenuti sono definiti da una query che fa riferimento a una o più tabelle. Puoi creare una vista che fa riferimento a un massimo di 10 tabelle utilizzando editor SQL per Amazon Athena o Amazon Redshift. Le tabelle di riferimento sottostanti per una vista possono appartenere allo stesso database o a database diversi all'interno dello stesso. Account AWS

Le seguenti considerazioni e limitazioni si applicano alle viste del catalogo dati.

- Amazon Redshift crea sempre viste con colonne varchar da tabelle con stringhe. È necessario trasmettere colonne di stringhe a varchar con una lunghezza esplicita quando si aggiungono dialetti da altri motori.
- La concessione delle autorizzazioni per il data lake all'`All views`interno di un database comporterà che il beneficiario disponga delle autorizzazioni su tutte le tabelle e le viste all'interno del database.
- Non puoi creare viste:

- Questo fa riferimento ad altre viste.
- Quando il riferimento a una tabella è un collegamento a una risorsa.
- Quando le tabelle di riferimento dispongono `IAM_ALLOWED_GROUP` delle autorizzazioni principali.
- Quando la tabella di riferimento si trova in un altro account.
- Da metastori Hive esterni.

Limitazioni del filtraggio dei dati

Quando concedi le autorizzazioni di Lake Formation su una tabella Data Catalog, puoi includere specifiche di filtraggio dei dati per limitare l'accesso a determinati dati nei risultati delle query e nei motori integrati con Lake Formation. Lake Formation utilizza il filtraggio dei dati per ottenere la sicurezza a livello di colonna, la sicurezza a livello di riga e la sicurezza a livello di cella. Puoi definire e applicare filtri di dati sulle colonne nidificate se i dati di origine contengono strutture nidificate.

Tieni presente le seguenti note e restrizioni per il filtraggio a livello di riga e cella.

- La sicurezza a livello di cella non è supportata nelle colonne annidate.
- Tutte le espressioni supportate nelle colonne di primo livello sono supportate anche nelle colonne nidificate. Tuttavia, non è necessario fare riferimento ai campi nidificati nelle colonne di partizione quando si definiscono espressioni nidificate a livello di riga.
- La sicurezza a livello di cella è disponibile in tutte le regioni quando si utilizza la versione 3 del motore Athena o Amazon Redshift Spectrum. Per altri servizi, la sicurezza a livello di cella è disponibile solo nelle regioni menzionate in. Regioni supportate
- Le istruzioni `SELECT INTO` non sono supportate.
- I `array` tipi di `map` dati e non sono supportati nelle espressioni di filtro di riga. Il tipo di `struct` dati è supportato.
- Per eseguire operazioni di interrogazione su tabelle che utilizzano filtri a livello di riga e cella, è necessario utilizzare un gruppo di lavoro speciale chiamato. `AmazonAthenaLakeFormation` Per informazioni sui gruppi di lavoro in Athena, consulta Using Workgroups for Running Queries nella Amazon Athena User Guide.
- Non c'è limite al numero di filtri di dati che possono essere definiti su una tabella, ma esiste un limite di 100 `SELECT` autorizzazioni di filtro dei dati per un singolo principale su una tabella.

- Il numero massimo di filtri di dati che possono essere inclusi in una concessione su una tabella è 10.
- Per applicare un filtro dati con un'espressione di filtro di riga, è necessario disporre `SELECT` dell'opzione grant su tutte le colonne della tabella. Questa restrizione non si applica agli amministratori degli account esterni quando la concessione è stata concessa all'account esterno.
- Se un principale è membro di un gruppo e sia al principale che al gruppo vengono concesse le autorizzazioni su un sottoinsieme di righe, i permessi di riga effettivi del principale sono l'unione dei permessi del principale e dei permessi del gruppo.
- I seguenti nomi di colonna sono limitati in una tabella per il filtraggio a livello di riga e di cella:
  - ctid
  - oid
  - xmin
  - cmin
  - xmax
  - cmax
  - tableoide
  - inserire xid
  - elimina xid
  - importoide
  - ID univoco per gatti rossi
- Se si applica l'espressione di filtro composta da tutte le righe su una tabella contemporaneamente ad altre espressioni di filtro con predicati, l'espressione composta da tutte le righe prevarrà su tutte le altre espressioni di filtro.
- Quando le autorizzazioni su un sottoinsieme di righe vengono concesse a un AWS account esterno e l'amministratore del data lake dell'account esterno concede tali autorizzazioni a un principale in quell'account, il predicato di filtro effettivo del principale è l'intersezione tra il predicato dell'account e qualsiasi predicato concesso direttamente al principale.

  Ad esempio, se l'account dispone delle autorizzazioni di riga con il predicato `dept='hr'` e al principale è stata concessa separatamente l'autorizzazione per`country='us'`, l'account principale ha accesso solo alle righe con `dept='hr' e``country='us'`.

Per ulteriori informazioni sul filtraggio a livello di cella, vedere. Filtraggio dei dati e sicurezza a livello di cella in Lake Formation

Considerazioni e limitazioni della modalità di accesso ibrido

La modalità di accesso ibrido offre la flessibilità necessaria per abilitare selettivamente le autorizzazioni di Lake Formation per database e tabelle del tuo. AWS Glue Data Catalog Con la modalità di accesso ibrida, ora disponi di un percorso incrementale che ti consente di impostare le autorizzazioni di Lake Formation per un set specifico di utenti senza interrompere le politiche di autorizzazione di altri utenti o carichi di lavoro esistenti.

Le seguenti considerazioni e limitazioni si applicano alla modalità di accesso ibrida.

Limitazioni

- Aggiorna la registrazione delle sedi Amazon S3: non puoi modificare i parametri di una sede registrata con Lake Formation utilizzando un ruolo collegato al servizio.
- Opzione di attivazione quando si utilizzano i tag LF — Quando è possibile concedere i permessi di Lake Formation utilizzando i tag LF, è possibile attivare i principali per applicare i permessi di Lake Formation in un passaggio consecutivo scegliendo database e tabelle a cui sono allegati i tag LF.
- Principi di opt-in: attualmente, solo un ruolo di amministratore del data lake può aggiungere i principali alle risorse.
- Attivazione di tutte le tabelle di un database: nelle concessioni tra account, quando si concedono autorizzazioni e in tutte le tabelle di un database, è necessario attivare anche il database affinché le autorizzazioni funzionino.

Considerazioni

- Aggiornamento della posizione Amazon S3 registrata con Lake Formation alla modalità di accesso ibrida: non consigliamo di convertire una posizione dati Amazon S3 già registrata con Lake Formation in modalità di accesso ibrida, sebbene sia possibile farlo.
- Comportamenti delle API quando una posizione dei dati viene registrata in modalità di accesso ibrida
  - CreateTable — La località è considerata registrata presso Lake Formation indipendentemente dal flag della modalità di accesso ibrida e dallo stato di attivazione. Pertanto, l'utente richiede l'autorizzazione alla localizzazione dei dati per creare una tabella.

- CreatePartition/BatchCreatePartitions/UpdatePartitions (quando la posizione della partizione viene aggiornata in modo che punti alla posizione registrata con hybrid): la posizione Amazon S3 è considerata registrata presso Lake Formation indipendentemente dal flag della modalità di accesso ibrida e dallo stato di attivazione. Pertanto, l'utente richiede l'autorizzazione alla localizzazione dei dati per creare o aggiornare un database.
- CreateDatabase/UpdateDatabase (quando la posizione del database viene aggiornata in modo che punti alla posizione registrata in modalità di accesso ibrida): la posizione viene considerata registrata con Lake Formation indipendentemente dal flag della modalità di accesso ibrida e dallo stato di attivazione. Pertanto, l'utente richiede l'autorizzazione alla localizzazione dei dati per creare o aggiornare un database.
- UpdateTable (quando una posizione della tabella viene aggiornata in modo che punti alla posizione registrata in modalità di accesso ibrida): la posizione viene considerata registrata con Lake Formation indipendentemente dal flag della modalità di accesso ibrida e dallo stato di attivazione. Pertanto, l'utente richiede l'autorizzazione alla localizzazione dei dati per aggiornare la tabella. Se la posizione della tabella non è aggiornata o punta a una posizione non registrata con Lake Formation, l'utente non richiede l'autorizzazione alla posizione dei dati per aggiornare la tabella.

Considerazioni e limitazioni sulla condivisione dei dati dei metadati di Hive

Con la federazione dei AWS Glue Data Catalog metadati (Data Catalog federation), puoi connettere il Data Catalog a metastore esterni che archiviano i metadati per i tuoi dati Amazon S3 e gestire in modo sicuro le autorizzazioni di accesso ai dati utilizzando. AWS Lake Formation

Le seguenti considerazioni e limitazioni si applicano ai database federati creati dai database Hive:

Considerazioni

- AWS SAM supporto delle applicazioni: sei responsabile della disponibilità delle risorse applicative che AWS SAM distribuisce (Amazon API Gateway e della funzione Lambda). Assicurati che la connessione tra il metastore AWS Glue Data Catalog e Hive funzioni quando gli utenti eseguono le query.
- Requisiti della versione di Hive metastore: puoi creare database federati solo utilizzando Apache Hive versione 3 e successive.

- Requisiti del database mappato: ogni database Hive deve essere mappato su un nuovo database in Lake Formation.
- Supporto federativo a livello di database: puoi connetterti a Hive metastore solo a livello di database.
- Autorizzazioni su database federati: le autorizzazioni applicate a un database federato o alle tabelle in un database federato persistono anche quando viene eliminato un database o una tabella di origine. Quando il database o la tabella di origine vengono ricreati, non è necessario concedere nuovamente le autorizzazioni. Quando una tabella federata con autorizzazioni Lake Formation viene eliminata all'origine, le autorizzazioni di Lake Formation sono ancora visibili e puoi revocarle se necessario.

  Se un utente elimina un database federato, tutte le autorizzazioni corrispondenti vengono perse. Ricreando lo stesso database con lo stesso nome, non verranno ripristinate le autorizzazioni di Lake Formation. Gli utenti dovranno configurare nuovamente le nuove autorizzazioni.
- Autorizzazioni di AllowedPrincipal gruppo IAM su database federati: in base a`DataLakeSettings`, Lake Formation potrebbe impostare le autorizzazioni per tutti i database e le tabelle a un gruppo virtuale denominato. `IAMAllowedPrincipal` `IAMAllowedPrincipal`Si riferisce a tutti i dirigenti IAM che hanno accesso alle risorse del Data Catalog tramite le politiche principali e le politiche delle risorse IAM. AWS Glue Se queste autorizzazioni esistono su un database o una tabella, a tutti i principali viene concesso l'accesso al database o alla tabella.

  Tuttavia, Lake Formation non consente `IAMAllowedPrincipal` le autorizzazioni sulle tabelle nei database federati. Quando crei database federati, assicurati di passare il `CreateTableDefaultPermissions` parametro come elenco vuoto.

  Per ulteriori informazioni, consulta Modifica delle impostazioni predefinite per il data lake.
- Unire tabelle nelle query: puoi unire le tabelle metastore di Hive alle tabelle native di Data Catalog per eseguire le query.

Limitazioni

- Limitazione alla sincronizzazione dei metadati tra il metastore Hive AWS Glue Data Catalog e il metastore Hive: dopo aver stabilito la connessione al metastore Hive, è necessario creare un database federato per sincronizzare i metadati nel metastore Hive con il. AWS Glue Data Catalog Le tabelle del database federato vengono sincronizzate in fase di esecuzione quando gli utenti eseguono le query.

- Limitazione alla creazione di nuove tabelle in un database federato: non sarà possibile creare nuove tabelle in database federati.
- Limitazione delle autorizzazioni dei dati: il supporto per le autorizzazioni sulle viste delle tabelle dei metastore di Hive non è disponibile.

Limitazioni della condivisione dei dati di Amazon Redshift

AWS Lake Formation consente di gestire in modo sicuro i dati in un datashare di Amazon Redshift. Amazon Redshift è un servizio di data warehouse completamente gestito su scala di petabyte nel cloud. AWS Utilizzando la funzionalità di condivisione dei dati, Amazon Redshift ti aiuta a condividere i dati tra di loro. Account AWS Per ulteriori informazioni sulla condivisione dei dati di Amazon Redshift, consulta Panoramica sulla condivisione dei dati in Amazon Redshift.

Le seguenti note e restrizioni si applicano ai database federati creati da condivisioni di dati Amazon Redshift:

- Requisiti del database mappato: ogni datashare Amazon Redshift deve essere mappato su un nuovo database in Lake Formation. Ciò è necessario per mantenere nomi di tabella univoci quando la rappresentazione degli oggetti datashare è appiattita nel database Data Catalog.
- Limitazione alla creazione di nuove tabelle in un database federato: non sarà possibile creare nuove tabelle in database federati.
- Autorizzazioni sui database federati: le autorizzazioni applicate a un database federato o alle tabelle in un database federato persistono anche quando viene eliminato una tabella di origine o un database. Quando il database o la tabella di origine vengono ricreati, non è necessario concedere nuovamente le autorizzazioni. Quando una tabella federata con autorizzazioni Lake Formation viene eliminata all'origine, le autorizzazioni di Lake Formation saranno ancora visibili e potrai revocarle se necessario.

  Se un utente elimina un database federato, tutte le autorizzazioni corrispondenti vengono perse. Ricreando lo stesso database con lo stesso nome, non verranno ripristinate le autorizzazioni di Lake Formation. Gli utenti dovranno configurare nuovamente le nuove autorizzazioni.
- Autorizzazioni di AllowedPrincipal gruppo IAM su database federati: in base a`DataLakeSettings`, Lake Formation potrebbe impostare le autorizzazioni per tutti i database e le tabelle a un gruppo virtuale denominato. `IAMAllowedPrincipal` `IAMAllowedPrincipal`Si riferisce a tutti i dirigenti IAM che hanno accesso alle risorse del Data Catalog tramite le politiche

principali e le politiche delle risorse IAM. AWS Glue Se queste autorizzazioni esistono su un database o una tabella, a tutti i principali viene concesso l'accesso al database o alla tabella.

Tuttavia, Lake Formation non consente `IAMAllowedPrincipal` le autorizzazioni sulle tabelle nei database federati. Quando crei database federati, assicurati di passare il `CreateTableDefaultPermissions` parametro come elenco vuoto.

Per ulteriori informazioni, consulta Modifica delle impostazioni predefinite per il data lake.

- Filtraggio dei dati: in Lake Formation, puoi concedere autorizzazioni su una tabella in un database federato con filtri a livello di colonna e di riga. Tuttavia, non è possibile combinare il filtraggio a livello di colonna e a livello di riga per limitare l'accesso con granularità a livello di cella alle tabelle all'interno di database federati.
- Identificatore di distinzione tra maiuscole e minuscole: gli oggetti datashare di Amazon Redshift gestiti da Lake Formation supporteranno i nomi delle tabelle e delle colonne solo in lettere minuscole. Non attivare l'identificatore di distinzione tra maiuscole e minuscole per database, tabelle e colonne nelle condivisioni di dati Amazon Redshift, se verranno condivise e gestite utilizzando Lake Formation.

Per ulteriori informazioni sulle limitazioni quando si lavora con le condivisioni di dati in Amazon Redshift, consulta la sezione Limitazioni per la condivisione dei dati nella Amazon Redshift Database Developer Guide.

Limitazioni dell'integrazione di IAM Identity Center

Con AWS IAM Identity Center, puoi connetterti ai provider di identità (IdPs) e gestire centralmente l'accesso per utenti e gruppi a tutti i servizi di AWS analisi. Puoi configurarlo AWS Lake Formation come applicazione abilitata in IAM Identity Center e gli amministratori del data lake possono concedere autorizzazioni granulari a utenti e gruppi autorizzati sulle risorse. AWS Glue Data Catalog

Le seguenti limitazioni si applicano all'integrazione di Lake Formation con IAM Identity Center:

- Non puoi assegnare utenti e gruppi di IAM Identity Center come amministratori di data lake o amministratori di sola lettura in Lake Formation.
- Gli utenti e i gruppi di IAM Identity Center non possono interrogare le tabelle del Data Catalog crittografate utilizzando le chiavi (). AWS Key Management Service AWS KMS AWS KMS non supporta la propagazione affidabile delle identità.

- Gli utenti e i gruppi di IAM Identity Center possono richiamare solo le operazioni API elencate nella `AWSIAMIdentityCenterAllowListForIdentityContext` policy fornita da IAM Identity Center.

Buone pratiche e considerazioni per il controllo degli accessi basato su tag Lake Formation

È possibile creare, gestire e assegnare LF-tags per controllare l'accesso ai database, alle tabelle e alle colonne del Data Catalog.

Prendi in considerazione le seguenti best practice quando utilizzi il controllo degli accessi basato su tag Lake Formation:

- Tutti i tag LF devono essere predefiniti prima di poter essere assegnati alle risorse del Data Catalog o concessi ai responsabili.

  L'amministratore del data lake può delegare le attività di gestione dei tag creando creatori di tag LF con le autorizzazioni IAM richieste. Gli ingegneri e gli analisti dei dati decidono le caratteristiche e le relazioni dei tag LF. I creatori di LF-tag creano quindi e mantengono i tag LF in Lake Formation.
- È possibile assegnare più LF-tag alle risorse del Data Catalog. È possibile assegnare un solo valore per una particolare chiave a una particolare risorsa.

  Ad esempio, è possibile assegnare`module=Orders`, `region=Westdivision=Consumer`, e così via a un database, una tabella o una colonna. Non puoi `module=Orders,Customers` assegnare.
- Non è possibile assegnare tag LF alle risorse quando si crea la risorsa. È possibile aggiungere tag LF solo alle risorse esistenti.
- È possibile concedere espressioni LF-Tag, non solo singoli tag LF, a un principale.

  Un'espressione LF-Tag ha un aspetto simile alla seguente (in pseudo-codice).

  ```
  module=sales AND division=(consumer OR commercial)
  ```

  Un principale a cui è concessa questa espressione LF-Tag può accedere solo alle risorse del Data Catalog (database, tabelle e colonne) assegnate e a una delle due opzioni. *`module=sales`* `division=consumer` `division=commercial` Se desideri che il principale sia in grado di accedere a risorse che dispongono `module=sales` o meno `division=commercial` di

includerle entrambe nella stessa concessione. Fai due sovvenzioni, una per `module=sales` e una per`division=commercial`.

L'espressione più semplice del tag LF è costituita da un solo tag LF, ad esempio. `module=sales`

- Un principale a cui sono concesse le autorizzazioni per un tag LF con più valori può accedere alle risorse del Data Catalog con uno di questi valori. Ad esempio, se a un utente viene concesso un tag LF con key= `module` e values=`orders,customers`, l'utente ha accesso alle risorse assegnate o. `module=orders` `module=customers`
- È necessario disporre dell'`Grant with LF-Tag expressions`autorizzazione per concedere le autorizzazioni relative ai dati sulle risorse del Catalogo dati utilizzando il metodo LF-TBAC. L'amministratore del data lake e il creatore di LF-Tag ricevono implicitamente questa autorizzazione. Un principale che dispone dell'`Grant with LFTag expressions`autorizzazione può concedere le autorizzazioni relative ai dati sulle risorse utilizzando:
  - il metodo di risorsa denominato
  - il metodo LF-TBAC, ma utilizzando solo la stessa espressione LF-TAG

  Ad esempio, supponiamo che l'amministratore del data lake conceda la seguente concessione (in pseudo-codice).

  ```
  GRANT (SELECT ON TABLES) ON TAGS module=customers, region=west,south TO user1 WITH
   GRANT OPTION
  ```

  In questo caso, `user1` può concedere `SELECT` tabelle ad altri principali utilizzando il metodo LF-TBAC, ma solo con l'espressione LF-TAG completa. `module=customers, region=west,south`
- Se a un principale vengono concesse le autorizzazioni su una risorsa sia con il metodo LF-TBAC che con il metodo della risorsa denominata, i permessi che il principale ha sulla risorsa sono l'unione dei permessi concessi con entrambi i metodi.
- Lake Formation supporta la concessione `DESCRIBE` e `ASSOCIATE` l'eliminazione dei tag LF tra gli account e la concessione di autorizzazioni sulle risorse del Data Catalog tra gli account utilizzando il metodo LF-TBAC. In entrambi i casi, il principale è l'ID dell'account. AWS

> **Note**
>
> Lake Formation supporta sovvenzioni tra account a organizzazioni e unità organizzative utilizzando il metodo LF-TBAC. Per utilizzare questa funzionalità, è necessario aggiornare le impostazioni della versione di Cross account alla versione 3.

Per ulteriori informazioni, consulta Condivisione dei dati tra account in Lake Formation.

- Le risorse del Data Catalog create in un account possono essere etichettate solo utilizzando i tag LF creati nello stesso account. I tag LF creati in un account non possono essere associati a risorse condivise di un altro account.
- L'utilizzo del controllo degli accessi basato su tag Lake Formation (LF-TBAC) per concedere l'accesso tra account alle risorse del Data Catalog richiede aggiunte alla politica delle risorse del Data Catalog per il tuo account. AWS Per ulteriori informazioni, consulta Prerequisiti.
- Le chiavi LF-Tag e i valori LF-Tag non possono superare i 50 caratteri di lunghezza.
- Il numero massimo di tag LF che possono essere assegnati a una risorsa del Data Catalog è 50.
- I seguenti limiti sono limiti flessibili:
  - Il numero massimo di tag LF che è possibile creare è 1000.
  - Il numero massimo di valori che possono essere definiti per un LF-tag è 1000.
- Le chiavi e i valori dei tag vengono convertiti in lettere minuscole quando vengono memorizzati.
- È possibile assegnare un solo valore per un tag LF a una particolare risorsa.
- Se vengono concessi più LF-tag a un principale con un'unica concessione, il principale può accedere solo alle risorse del Data Catalog che contengono tutti i tag LF.
- AWS Glue I lavori ETL richiedono l'accesso completo alla tabella. I processi falliranno se il ruolo AWS Glue ETL non ha accesso a tutte le colonne di una tabella. È possibile applicare i tag LF a livello di colonna, ma ciò può far sì che i ruoli AWS Glue ETL perdano l'accesso completo alla tabella e che i job falliscano. L'utilizzo di filtri di dati per il filtraggio di colonne e/o righe non è interessato da questa limitazione.
- Se la valutazione di un'espressione LF-Tag comporta l'accesso solo a un sottoinsieme di colonne della tabella, ma l'autorizzazione Lake Formation concessa in caso di corrispondenza è una delle autorizzazioni che richiedevano l'accesso completo alla colonna, vale a dire,, o `Alter` `Drop` `InsertDelete`, allora nessuna di queste autorizzazioni viene concessa. Invece, viene concesso

solo. `Describe` Se l'autorizzazione concessa è `All` (`Super`), allora solo `Select` e `Describe` vengono concesse.

- I wildcard non vengono utilizzati con i tag LF. Per assegnare un tag LF a tutte le colonne di una tabella, si assegna il tag LF alla tabella e tutte le colonne della tabella ereditano il tag LF. Per assegnare un tag LF a tutte le tabelle di un database, si assegna il tag LF al database e tutte le tabelle del database ereditano quel tag LF.

Formati e limitazioni supportati per la compattazione gestita dei dati

Per migliorare le prestazioni di lettura tramite servizi di AWS analisi come Amazon Athena, Amazon EMR ed AWS Glue ETL Jobs, AWS Glue Data Catalog offre la compattazione gestita (un processo che compatta piccoli oggetti Amazon S3 in oggetti più grandi) per le tabelle Iceberg in Data Catalog.

La compattazione dei dati supporta una varietà di tipi di dati e formati di compressione per la lettura e la scrittura di dati, inclusa la lettura di dati da tabelle crittografate.

La compattazione dei dati supporta:

- Tipi di dati: Boolean, Integer, Lungo, Float, Doppio, Stringa, Decimale, Data, Ora, Timestamp, Stringa, UUID, Binary
- Compressione: zstd, gzip, snappy, non compresso
- Crittografia: la compattazione dei dati supporta solo la crittografia Amazon S3 (SSE-S3) e la crittografia KMS lato server (SSE-KMS).
- Compattazione BinPack
- Evoluzione dello schema
- Tabelle con dimensione del file di destinazione (scrittura). target-file-size-bytes proprietà (in configurazione iceberg) nell'intervallo compreso tra 128 MB e 512 MB.
- Regioni
  - Asia Pacifico (Tokyo)
  - Asia Pacifico (Seul)
  - Asia Pacifico (Mumbai)
  - Europa (Irlanda)
  - Europa (Francoforte)
  - Stati Uniti orientali (Virginia settentrionale)

  - Stati Uniti orientali (Ohio)
  - Stati Uniti occidentali (California settentrionale)
  - Sud America (San Paolo)
- Puoi eseguire la compattazione dall'account in cui risiede il Catalogo dati quando il bucket Amazon S3 che archivia i dati sottostanti si trova in un altro account. Per eseguire questa azione, il ruolo di compattazione richiede l'accesso al bucket Amazon S3.

La compattazione dei dati attualmente non supporta:

- Tipi di dati: fissi
- Compressione: brotli, lz4
- Compattazione dei file mentre le specifiche della partizione si evolvono.
- Ordinamento regolare o con ordine z
- Unisci o elimina file: il processo di compattazione non considera i file di dati a cui sono associati file eliminati.
- Compattazione su tabelle con più account: non è possibile eseguire la compattazione su tabelle con più account.
- Compattazione su tabelle interregionali: non è possibile eseguire la compattazione su tabelle interregionali.
- Abilitazione della compattazione sui link alle risorse
- Endpoint VPC per bucket Amazon S3

Risoluzione dei problemi relativi a Lake Formation

Se riscontri problemi quando lavori con AWS Lake Formation, consulta gli argomenti di questa sezione.

Argomenti

Risoluzione dei problemi generali

Usa le informazioni qui per aiutarti a diagnosticare e risolvere vari problemi di Lake Formation.

Errore: autorizzazioni Lake Formation insufficienti su <Amazon S3 location>

È stato effettuato un tentativo di creare o modificare una risorsa Data Catalog senza autorizzazioni di localizzazione dei dati sulla posizione Amazon S3 indicata dalla risorsa.

Se un database o una tabella Data Catalog punta a una posizione Amazon S3, quando concedi le autorizzazioni a Lake Formation `CREATE_TABLE` oppure `ALTER` devi anche concedere l'`DATA_LOCATION_ACCESS`autorizzazione sulla posizione. Se concedi queste autorizzazioni ad account esterni o a organizzazioni, devi includere l'opzione di concessione.

Dopo aver concesso queste autorizzazioni a un account esterno, l'amministratore del data lake di quell'account deve quindi concedere le autorizzazioni ai principali (utenti o ruoli) dell'account. Quando concedi l'`DATA_LOCATION_ACCESS`autorizzazione ricevuta da un altro account, devi specificare l'ID di catalogo (ID account) dell'AWSaccount proprietario. L'account del proprietario è l'account che ha registrato la sede.

Per ulteriori informazioni, consultare Controllo sottostante dell'accesso ai dati e Concessione delle autorizzazioni per la localizzazione dei dati.

Errore: «Autorizzazioni per la chiave di crittografia insufficienti per l'API Glue»

È stato fatto un tentativo di concedere autorizzazioni a Lake Formation senza autorizzazioni AWS Identity and Access Management (IAM) sulla chiave di AWS KMS crittografia per un Data Catalog crittografato.

La mia query Amazon Athena o quella di Amazon Redshift che utilizza i manifesti non riesce

Lake Formation non supporta le query che utilizzano i manifest.

Errore: «Autorizzazioni di Lake Formation insufficienti: è richiesta la creazione di un tag sul catalogo»

L'utente/ruolo deve essere un amministratore del data lake.

Errore durante l'eliminazione di amministratori di data lake non validi

È necessario eliminare contemporaneamente tutti gli amministratori di data lake non validi (ruoli IAM eliminati definiti come amministratori di data lake). Se si tenta di eliminare separatamente gli amministratori di data lake non validi, Lake Formation genera un errore principale non valido.

Risoluzione dei problemi di accesso tra account

Usa le informazioni qui per aiutarti a diagnosticare e risolvere i problemi di accesso su più account.

Argomenti

- Errore: «Accesso negato per recuperare le informazioni AWS sull'organizzazione»
- Errore: «Organizzazione <organization-ID>non trovata»
- Errore: «Autorizzazioni Lake Formation insufficienti: combinazione non valida»
- ConcurrentModificationException sulle richieste di concessione/revoca ad account esterni
- Errore durante l'utilizzo di Amazon EMR per accedere ai dati condivisi tramite più account

Ho concesso un'autorizzazione Lake Formation per più account ma il destinatario non può vedere la risorsa

- L'utente dell'account del destinatario è un amministratore di Data Lake? Solo gli amministratori del data lake possono vedere la risorsa al momento della condivisione.
- Stai condividendo con un account esterno alla tua organizzazione utilizzando il metodo della risorsa denominata? In tal caso, l'amministratore del data lake dell'account destinatario deve accettare un invito alla condivisione delle risorse in AWS Resource Access Manager (AWS RAM).

  Per ulteriori informazioni, consulta the section called “Accettazione di un invito alla condivisioneAWS RAM delle risorse.”.
- Stai utilizzando politiche relative alle risorse a livello di account (Data Catalog) in? AWS Glue In caso affermativo, se si utilizza il metodo delle risorse denominate, è necessario includere una dichiarazione speciale nella politica che autorizzi AWS RAM a condividere le politiche per conto dell'utente.

  Per ulteriori informazioni, consulta the section called “Gestione delle autorizzazioni tra account utilizzando entrambi AWS Glue e Lake Formation”.
- Disponi delle autorizzazioni AWS Identity and Access Management (IAM) necessarie per concedere l'accesso a più account?

  Per ulteriori informazioni, consulta the section called “Prerequisiti”.
- La risorsa a cui hai concesso le autorizzazioni non deve avere alcuna autorizzazione Lake Formation concessa al `IAMAllowedPrincipals` gruppo.
- C'è una `deny` dichiarazione sulla risorsa nella politica a livello di account?

I responsabili dell'account del destinatario possono vedere la risorsa Data Catalog ma non possono accedere ai dati sottostanti

I principali dell'account del destinatario devono disporre delle autorizzazioni richieste AWS Identity and Access Management (IAM). Per informazioni dettagliate, vedi Accesso ai dati sottostanti di una tabella condivisa.

Errore: «Associazione non riuscita perché il chiamante non era autorizzato» quando si accetta un AWS RAM invito alla condivisione delle risorse

Dopo aver concesso l'accesso a una risorsa a un altro account, quando l'account ricevente tenta di accettare l'invito alla condivisione delle risorse, l'azione fallisce.

```
$ aws ram get-resource-share-associations --association-type PRINCIPAL --resource-
share-arns arn:aws:ram:aws-region:444444444444:resource-share/e1d1f4ba-xxxx-xxxx-xxxx-
xxxxxxxx5d8d
{
    "resourceShareAssociations": [
        {
            "resourceShareArn": "arn:aws:ram:aws-region:444444444444:resource-share/
e1d1f4ba-xxxx-xxxx-xxxx-xxxxxxxx5d8d
",
            "resourceShareName": "LakeFormation-MMCC0XQBH3Y",
            "associatedEntity": "5815803XXXXX",
            "associationType": "PRINCIPAL",
            "status": "FAILED",
            "statusMessage": "Association failed because the caller was not
 authorized.",
            "creationTime": "2021-07-12T02:20:10.267000+00:00",
            "lastUpdatedTime": "2021-07-12T02:20:51.830000+00:00",
            "external": true
        }
    ]
}
```

L'errore si verifica perché `glue:PutResourcePolicy` viene richiamato da AWS Glue quando l'account ricevente accetta l'invito alla condivisione delle risorse. Per risolvere il problema, consenti l'`glue:PutResourcePolicy`azione da parte del ruolo assunto utilizzato dall'account produttore/concedente.

Errore: «Non autorizzato a concedere le autorizzazioni per la risorsa»

È stato effettuato un tentativo di concedere autorizzazioni per più account su un database o una tabella di proprietà di un altro account. Quando un database o una tabella vengono condivisi con il tuo account, in qualità di amministratore del data lake, puoi concedere le autorizzazioni solo agli utenti del tuo account.

Errore: «Accesso negato per recuperare le informazioni AWS sull'organizzazione»

Il tuo account è un account di gestione di AWS Organizations e non disponi delle autorizzazioni necessarie per recuperare le informazioni sull'organizzazione, come le unità organizzative nell'account.

Per ulteriori informazioni, consulta Required permissions for cross-account grants.

Errore: «Organizzazione <organization-ID>non trovata»

È stato effettuato un tentativo di condividere una risorsa con un'organizzazione, ma la condivisione con le organizzazioni non è abilitata. Abilita la condivisione delle risorse con le organizzazioni.

Per ulteriori informazioni, consulta Abilitare la condivisione con le organizzazioni AWS nella Guida per l'utente di AWS RAM.

Errore: «Autorizzazioni Lake Formation insufficienti: combinazione non valida»

Un utente ha condiviso una risorsa Data Catalog mentre al `IAMAllowedPrincipals` gruppo venivano concesse le autorizzazioni di Lake Formation per la risorsa. L'utente deve revocare tutte le autorizzazioni di Lake Formation `IAMAllowedPrincipals` prima di condividere la risorsa.

ConcurrentModificationException sulle richieste di concessione/revoca ad account esterni

Quando gli utenti effettuano più richieste di autorizzazione simultanee di concessione e/o revoca per un principale sulle politiche LF-Tag, Lake Formation lancia. ConcurrentModificationException Gli utenti devono catturare l'eccezione e riprovare la richiesta grant/revoke non riuscita. Utilizza versioni batch delle operazioni`GrantPermissions`/`RevokePermissions`API BatchGrantPermissionse

BatchRevokePermissionsallevia questo problema in una certa misura riducendo il numero di richieste di concessione/revoca simultanee.

Errore durante l'utilizzo di Amazon EMR per accedere ai dati condivisi tramite più account

Quando utilizzi Amazon EMR per accedere ai dati condivisi con te da un altro account, alcune librerie Spark tenteranno di chiamare l'operazione API `Glue:GetUserDefinedFunctions`. Poiché le versioni 1 e 2 delle autorizzazioni gestite da AWS RAM non supportano questa operazione, viene visualizzato il seguente messaggio di errore:

```
"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"
```

Per risolvere questo errore, l'amministratore del data lake che ha creato la condivisione di risorse deve aggiornare le autorizzazioni gestite da AWS RAM collegate alla condivisione di risorse. La versione 3 delle autorizzazioni gestite da AWS RAM consente ai responsabili di eseguire l'operazione `glue:GetUserDefinedFunctions`.

Se crei una nuova condivisione di risorse, Lake Formation applica la versione più recente dell'autorizzazione gestita da AWS RAM per impostazione predefinita senza richiedere alcuna azione da parte tua. Per abilitare l'accesso ai dati tra account per le condivisioni di risorse esistenti, è necessario aggiornare le autorizzazioni gestite da AWS RAM alla versione 3.

Puoi visualizzare le autorizzazioni AWS RAM assegnate alle risorse condivise con te in AWS RAM. Le autorizzazioni seguenti sono incluse nella versione 3:

```
Databases
  AWSRAMPermissionGlueDatabaseReadWriteForCatalog
  AWSRAMPermissionGlueDatabaseReadWrite

Tables
  AWSRAMPermissionGlueTableReadWriteForCatalog
  AWSRAMPermissionGlueTableReadWriteForDatabase

AllTables
  AWSRAMPermissionGlueAllTablesReadWriteForCatalog
```

```
  AWSRAMPermissionGlueAllTablesReadWriteForDatabase
```

Aggiornamento della versione delle autorizzazioni gestite da AWS RAM delle condivisioni di risorse esistenti

L'utente (amministratore del data lake) può aggiornare le autorizzazioni gestite da AWS RAM a una versione più recente seguendo le istruzioni riportate nella Guida per l'utente di AWS RAM oppure revocare tutte le autorizzazioni esistenti per il tipo di risorsa e concederle nuovamente. Se revochi le autorizzazioni, AWS RAM elimina la condivisione di risorse AWS RAM associata al tipo di risorsa. Quando concedi nuovamente le autorizzazioni, AWS RAM crea nuove condivisioni di risorse collegando la versione più recente delle autorizzazioni gestite da AWS RAM.

Risoluzione dei problemi di blueprint e flussi di lavoro

Utilizza le informazioni riportate qui per aiutarti a diagnosticare e risolvere i problemi relativi ai blueprint e al flusso di lavoro.

Argomenti

<role-ARN>Il mio progetto non è riuscito con «User: <user-ARN>is not authorized to perform: iam: PassRole on resource:»

È stato effettuato un tentativo di creare un blueprint da parte di un utente che non dispone di autorizzazioni sufficienti per assegnare il ruolo scelto.

Aggiorna la policy IAM dell'utente per poter passare il ruolo o chiedi all'utente di scegliere un ruolo diverso con le autorizzazioni passrole richieste.

Per ulteriori informazioni, consulta the section called “Riferimento ai personaggi di Lake Formation e alle autorizzazioni IAM”.

Il mio flusso di lavoro non è riuscito con «User: <user-ARN>is not authorized to perform: iam: PassRole on resource:<role-ARN>»

Il ruolo che hai specificato per il flusso di lavoro non aveva una politica in linea che consentisse al ruolo di passare da solo.

Per ulteriori informazioni, consulta the section called “(Facoltativo) Crea un ruolo IAM per i flussi di lavoro”.

Un crawler del mio flusso di lavoro non è riuscito con il messaggio «La risorsa non esiste o il richiedente non è autorizzato ad accedere alle autorizzazioni richieste»

Una possibile causa è che il ruolo passato non disponeva di autorizzazioni sufficienti per creare una tabella nel database di destinazione. Concedi al ruolo l'`CREATE_TABLE`autorizzazione per il database.

Un crawler del mio flusso di lavoro non è riuscito con il messaggio «Si è verificato un errore (AccessDeniedException) durante la chiamata dell' CreateTable operazione...»

Una possibile causa è che il ruolo del workflow non disponeva delle autorizzazioni per la localizzazione dei dati sulla posizione di archiviazione di destinazione. Concedi le autorizzazioni per la localizzazione dei dati al ruolo.

Per ulteriori informazioni, consulta the section called “`DATA_LOCATION_ACCESS`”.

Problemi noti per AWS Lake Formation

Esamina questi problemi noti per AWS Lake Formation.

Argomenti

- Limitazione al filtraggio dei metadati delle tabelle

- Problema con la ridenominazione di una colonna esclusa
- Problema con l'eliminazione delle colonne nelle tabelle CSV
- Le partizioni delle tabelle devono essere aggiunte in un percorso comune
- Problema con la creazione di un database durante la creazione del flusso di lavoro
- Problema con l'eliminazione e la successiva creazione di un utente
- GetTablese le SearchTables API non aggiornano il valore del parametro IsRegisteredWithLakeFormation
- Le operazioni dell'API Data Catalog non aggiornano il valore del IsRegisteredWithLakeFormation parametro
- Le operazioni di Lake Formation non supportano AWS Glue Schema Registry

Limitazione al filtraggio dei metadati delle tabelle

AWS Lake Formationle autorizzazioni a livello di colonna possono essere utilizzate per limitare l'accesso a colonne specifiche di una tabella. Quando un utente recupera i metadati sulla tabella utilizzando la console o un'API come`glue:GetTable`, l'elenco delle colonne nell'oggetto tabella contiene solo i campi a cui ha accesso. È importante comprendere i limiti di questo filtraggio dei metadati.

Sebbene Lake Formation renda disponibili i metadati sulle autorizzazioni delle colonne ai servizi integrati, l'effettivo filtraggio delle colonne nelle risposte alle query è responsabilità del servizio integrato. I client Lake Formation che supportano il filtraggio a livello di colonna, tra cui Amazon Athena, Amazon Redshift Spectrum e Amazon EMR, filtrano i dati in base alle autorizzazioni delle colonne registrate con Lake Formation. Gli utenti non saranno in grado di leggere dati a cui non dovrebbero avere accesso. Attualmente, AWS Glue ETL non supporta il filtraggio delle colonne.

> **Note**
>
> I cluster EMR non sono gestiti completamente da. AWS Pertanto, è responsabilità degli amministratori EMR proteggere adeguatamente i cluster per evitare l'accesso non autorizzato ai dati.

Alcune applicazioni o formati potrebbero memorizzare metadati aggiuntivi, inclusi nomi e tipi di colonne, nella mappa come proprietà della `Parameters` tabella. Queste proprietà vengono restituite

non modificate e sono accessibili da qualsiasi utente con `SELECT` autorizzazione su qualsiasi colonna.

Ad esempio, Avro SerDe memorizza una rappresentazione JSON dello schema della tabella in una proprietà della tabella denominata`avro.schema.literal`, disponibile per tutti gli utenti con accesso alla tabella. Si consiglia di evitare di memorizzare informazioni riservate nelle proprietà delle tabelle e di tenere presente che gli utenti possono apprendere lo schema completo delle tabelle in formato Avro. Questa limitazione è specifica per i metadati relativi a una tabella.

AWS Lake Formationrimuove qualsiasi proprietà della tabella a partire da `spark.sql.sources.schema` quando risponde a una richiesta `glue:GetTable` o a una richiesta simile se il chiamante non dispone `SELECT` delle autorizzazioni su tutte le colonne della tabella. Ciò impedisce agli utenti di accedere a metadati aggiuntivi sulle tabelle create con Apache Spark. Se eseguite su Amazon EMR, le applicazioni Apache Spark possono ancora leggere queste tabelle, ma alcune ottimizzazioni potrebbero non essere applicate e i nomi delle colonne con distinzione tra maiuscole e minuscole non sono supportati. Se l'utente ha accesso a tutte le colonne della tabella, Lake Formation restituisce la tabella non modificata con tutte le proprietà della tabella.

Problema con la ridenominazione di una colonna esclusa

Se si utilizzano le autorizzazioni a livello di colonna per escludere una colonna e quindi rinominarla, la colonna non viene più esclusa dalle query, ad esempio. `SELECT *`

Problema con l'eliminazione delle colonne nelle tabelle CSV

Se crei una tabella del catalogo dati in formato CSV e poi elimini una colonna dallo schema, le query potrebbero restituire dati errati e le autorizzazioni a livello di colonna potrebbero non essere rispettate.

Soluzione alternativa: create invece una nuova tabella.

Le partizioni delle tabelle devono essere aggiunte in un percorso comune

Lake Formation si aspetta che tutte le partizioni di una tabella si trovino in un percorso comune impostato nel campo della posizione della tabella. Quando si utilizza il crawler per aggiungere partizioni a un catalogo, ciò funziona perfettamente. Tuttavia, se si aggiungono partizioni manualmente e queste partizioni non si trovano nella posizione impostata nella tabella principale, l'accesso ai dati non funziona.

Problema con la creazione di un database durante la creazione del flusso di lavoro

Quando crei un flusso di lavoro da un blueprint utilizzando la console Lake Formation, puoi creare il database di destinazione se non esiste. Quando lo fai, l'utente che ha effettuato l'accesso ottiene l'`CREATE_TABLE`autorizzazione sul database che viene creato. Tuttavia, il crawler generato dal flusso di lavoro assume il ruolo del flusso di lavoro quando tenta di creare una tabella. Questa operazione non riesce perché il ruolo non dispone dell'`CREATE_TABLE`autorizzazione per il database.

Soluzione alternativa: se si crea il database tramite la console durante la configurazione del flusso di lavoro, prima di eseguire il flusso di lavoro, è necessario concedere al ruolo associato al flusso di lavoro l'`CREATE_TABLE`autorizzazione sul database appena creato.

Problema con l'eliminazione e la successiva creazione di un utente

Lo scenario seguente genera permessi errati per Lake Formation restituiti da: `lakeformation:ListPermissions`

1. Crea un utente e concedi le autorizzazioni di Lake Formation.
2. Eliminare l'utente.
3. Ricrea l'utente con lo stesso nome.

`ListPermissions`restituisce due voci, una per il vecchio utente e una per il nuovo utente. Se si tenta di revocare le autorizzazioni concesse al vecchio utente, le autorizzazioni vengono revocate al nuovo utente.

`GetTables`e le **`SearchTables`** API non aggiornano il valore del parametro **`IsRegisteredWithLakeFormation`**

Esiste una limitazione nota in base alla quale le operazioni dell'API Data Catalog come `GetTables` e `SearchTables` non aggiornano il valore di e restituiscono il valore predefinito, che è falso. `IsRegisteredWithLakeFormation parameter` Si consiglia di utilizzare l'`GetTable`API per visualizzare il valore corretto per`IsRegisteredWithLakeFormation parameter`.

Le operazioni dell'API Data Catalog non aggiornano il valore del **`IsRegisteredWithLakeFormation`** parametro

È noto che le operazioni dell'API Data Catalog, come `GetTables` e `SearchTables` non, aggiornano il valore del `IsRegisteredWithLakeFormation` parametro e restituiscono il valore predefinito, che è falso. Si consiglia di utilizzare l'`GetTable`API per visualizzare il valore corretto del `IsRegisteredWithLakeFormation` parametro.

Le operazioni di Lake Formation non supportano AWS Glue Schema Registry

Le operazioni di Lake Formation non supportano le AWS Glue tabelle che contengono un `SchemaReference` `StorageDescriptor` da utilizzare nello Schema Registry.

Messaggio di errore aggiornato

AWSLake Formation ha aggiornato le eccezioni specifiche delle risorse al messaggio di `EntityNotFound` errore generale per le seguenti operazioni API per soddisfare gli obiettivi di sicurezza e conformità.

- RevokePermissions
- GrantPermissions
- GetResourceTag LF
- GetTable
- GetDatabase

API AWS Lake Formation

> **Note**
>
> Il riferimento API aggiornato per il AWS Lake Formation servizio è ora disponibile.

Indice

- Modelli di stringa

API per le autorizzazioni

La sezione API delle autorizzazioni descrive le operazioni e i tipi di dati necessari per concedere e revocare le autorizzazioni inAWS Lake Formation. Consulta la Guida di riferimento alle API di Lake Formation per tutte le operazioni e i tipi di dati delleAWS Lake Formation API.

Operazioni

- GrantPermissions
- RevokePermissions
- BatchGrantPermissions
- BatchRevokePermissions
- GetEffectivePermissionsForPath
- ListPermissions

Tipi di dati

- Resource (Risorsa)
- DatabaseResource
- TableResource
- TableWithColumnsResource
- DataCellsFilterResourcee
- DataLocationResource
- DataLakePrincipal
- PrincipalPermissions
- PrincipalResourcePermissions
- DetailsMap
- ColumnWildcard
- BatchPermissionsRequestEntry
- BatchPermissionsFailureEntry

API per le impostazioni del data lake

Questa sezione contiene le impostazioni del data lake, le operazioni API e i tipi di dati per la gestione degli amministratori del data lake.

Operazioni

- GetDataLakeSettings
- PutDataLakeSettings

Tipi di dati

- DataLakeSettings

API di integrazione IAM Identity Center

Questa sezione contiene le operazioni per creare e gestire l'integrazione di Lake Formation con IAM Identity Center.

Operazioni

- CreateLakeFormationIdentityCenterConfiguration
- DeleteLakeFormationIdentityCenterConfiguration
- DescribeLakeFormationIdentityCenterConfiguration
- UpdateLakeFormationIdentityCenterConfiguration

Tipi di dati

- ExternalFilteringConfiguration

API in modalità di accesso ibrido

La sezione API della modalità di accesso ibrida descrive le operazioni e i tipi di dati necessari per configurare la modalità di accesso ibrida inAWS Lake Formation. Consulta la Guida di riferimento dell'API Lake Formation per tutte le operazioni e i tipi di dati delle AWS Lake Formation API.

Operazioni

- CreateLakeFormationOptIn
- DeleteLakeFormationOptIn
- ListLakeFormationOptIns

Tipi di dati

- Resource (Risorsa)
- DatabaseResource
- TableResource
- Informazioni sulle risorse
- LakeFormationOptInsInfo
- DataLocationResource

API di vendita di credenziali

La sezione Credential Vending API descrive le operazioni e i tipi di dati relativi all'utilizzo del AWS Lake Formation servizio per la fornitura di credenziali e la registrazione e la gestione di una risorsa data lake.

Operazioni

- RegisterResource
- DeregisterResource
- ListResources
- GetUnfilteredTableMetadata
- GetUnfilteredPartitionsMetadata
- GetTemporaryGluePartitionCredentials
- GetTemporaryGlueTableCredentials
- UpdateResource

Tipi di dati

- FilterCondition
- RowFilter
- ResourceInfo

API per l'assegnazione di tag

La sezione Tagging API descrive le operazioni e i tipi di dati relativi a una strategia di autorizzazione che definisce un modello di autorizzazioni sugli attributi o sui tag della coppia chiave-valore.

Operazioni

- Aggiungi LFTagsToResource
- Rimuovi LFTagsFromResource
- GetResourceEtichette LF
- Elenco dei tag LF
- Crea tag LF
- Ottieni il tag LF
- Aggiorna tag LF
- Elimina il tag ELF
- SearchTablesByEtichette LF
- SearchDatabasesByEtichette LF

Tipi di dati

- LFTagKeyResource
- LFTagPolicyResource
- TaggedTable
- TaggedDatabase
- Etichetta LF
- LFTagPair

- LFTagError
- Etichetta LF a colonna

API di filtraggio dei dati

Le API Data Filter descrivono come gestire i filtri delle celle di dati inAWS Lake Formation.

Operazioni

- CreateDataCellsFilter
- DeleteDataCellsFilter
- ListDataCellsFilter
- GetDataCellsFilter
- UpdateDataCellsFilter

Tipi di dati

- DataCellsFilter
- RowFilter

Tipi di dati comuni

Common Data Types descrive i vari tipi di dati comuni in AWS Lake Formation.

ErrorDetail struttura

Contiene dettagli su un errore.

Campi

- `ErrorCode`: stringa UTF-8, non inferiore a 1 o superiore a 255 byte di lunghezza, corrispondente a Single-line string pattern.

  Il codice associato a questo errore.
- `ErrorMessage`: stringa di descrizione, non superiore a 2048 byte di lunghezza, corrispondente a URI address multi-line string pattern.

Messaggio che descrive l'errore.

Modelli di stringa

L'API usa le seguenti espressioni regolari per definire i contenuti validi per vari membri e parametri di stringa:

- Modello di stringa a una riga: "[\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\t]*"
- IndirizzoModello di stringa a più righe per indirizzo URI: "[\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]*"
- Modello di stringhe personalizzato #3 — "^\w+\.\w+\.\w+$»
- Modello di stringhe personalizzato #4 — "^\w+\.\w+$»
- Modello di stringhe personalizzato #5 — "arn:aws:iam::[0-9]*:role/.*»
- Modello di stringhe personalizzato #6 — "arn:aws:iam::[0-9]*:user/.*»
- Modello di stringhe personalizzato #7 — "arn:aws:iam::[0-9]*:group/.*»
- Pattern di stringa personalizzato n. 8 – "arn:aws:iam::[0-9]*:saml-provider/.*"
- Pattern di stringa personalizzato n. 9 – "^([\p{L}\p{Z}\p{N}_.:\/=+\-@%]*)$"
- Pattern di stringa personalizzato n #10: "^([\p{L}\p{Z}\p{N}_.:*\/=+\-@%]*)$"
- Pattern di stringa personalizzato n #11: "[\p{L}\p{N}\p{P}]*"

Regioni supportate

Questa sezione contiene informazioni sul supporto Regioni AWS e sulle funzionalità di Lake Formation.

Disponibilità generale

Per i servizi Regioni AWS supportati da AWS Lake Formation, consulta Elenco dei AWS servizi disponibili per regione.

Per un elenco degli endpoint del servizio Lake Formation per ogni regione e delle quote del servizio Lake Formation, vedi AWS Lake Formation endpoint e quote.

AWS GovCloud (US)

Per una panoramica delle differenze tra AWS GovCloud (US) Region e standard Regioni AWS, vedi How differences for. AWS Lake FormationAWS GovCloud (US)

Ottimizzazione delle transazioni e dello storage

Le tabelle gestite, il supporto delle transazioni e le funzionalità di ottimizzazione dello storage per Lake Formation sono disponibili nelle seguenti Regioni AWS versioni:

Nome Regione	Parametro della regione	Endpoint
US East (N. Virginia)	us-east-1	lakeformation.us-east-1.amazonaws.com lakeformation-fips.us-east-1.amazonaws.com
Stati Uniti orientali (Ohio)	us-east-2	lakeformation.us-east-2.amazonaws.com lakeformation-fips.us-east-2.amazonaws.com

Nome Regione	Parametro della regione	Endpoint
US West (Oregon)	us-west-2	lakeformation.us-west-2.amazonaws.com lakeformation-fips.us-west-2.amazonaws.com
Asia Pacific (Mumbai)	ap-south-1	lakeformation.ap-south-1.amazonaws.com
Asia Pacifico (Seoul)	ap-northeast-2	lakeformation.ap-northeast-2.amazonaws.com
Asia Pacific (Singapore)	ap-southeast-1	lakeformation.ap-southeast-1.amazonaws.com
Asia Pacific (Sydney)	ap-southeast-2	lakeformation.ap-southeast-2.amazonaws.com
Asia Pacifico (Tokyo)	ap-northeast-1	lakeformation.ap-northeast-1.amazonaws.com
Europe (Frankfurt)	eu-central-1	lakeformation.eu-central-1.amazonaws.com
Europa (Irlanda)	eu-west-1	lakeformation.eu-west-1.amazonaws.com
Europe (London)	eu-west-2	lakeformation.eu-west-2.amazonaws.com
Europa (Stoccolma)	eu-north-1	lakeformation.eu-north-1.amazonaws.com

Nome Regione	Parametro della regione	Endpoint
Canada (Central)	ca-central-1	lakeformation.ca-central-1.amazonaws.com
Sud America (São Paulo)	sa-east-1	lakeformation.sa-east-1.amazonaws.com

Cronologia dei documenti per AWS Lake Formation

Nella tabella seguente vengono descritte importanti modifiche alla documentazione di AWS Lake Formation.

Modifica	Descrizione	Data
Configurazione aggiornata di Lake Formation	Sono stati aggiornati i passaggi nella AWS Lake Formation sezione Configurazione.	7 febbraio 2024
Modifica aggiornata della politica	Sono state aggiunte nuove autorizzazioni alla politica in linea del ruolo collegato al servizio. Per ulteriori informazioni, consulta Using service-linked roles for Lake Formation.	7 febbraio 2024
Modifica aggiornata della politica	Ha documentato la modifica alla LakeFormationDataAccessServiceRolePolicypolitica.	2 febbraio 2024
Limitazioni di Consolidated Lake Formation	Creata una sezione unificata per i limiti e le considerazioni di Lake Formation. Per ulteriori informazioni, consulta le limitazioni di Lake Formation.	15 dicembre 2023
Documentazione aggiunta per la compattazione di Iceberg	Per migliorare le prestazioni di lettura, servizi di AWS analisi come Athena e Amazon EMR e processi AWS Glue ETL, AWS Glue Data Catalog offre la compattazione gestita (un	25 novembre 2023

	processo che compatta piccoli oggetti Amazon S3 in oggetti più grandi) per le tabelle Iceberg nel Data Catalog. Per ulteriori informazioni, consulta Ottimizzazione delle tabelle Iceberg.	
Documentazione aggiunta per l'integrazione con IAM Identity Center	Le integrazioni di IAM Identity Center consentono a utenti e gruppi di accedere alle risorse del Data Catalog applicando le autorizzazioni di Lake Formation. Per ulteriori informazioni, consulta l'integrazione con IAM Identity Center.	25 novembre 2023
È stata aggiunta documentazione per le visualizzazioni del catalogo dati	Puoi creare viste AWS Glue Data Catalog che fanno riferimento a un massimo di 10 tabelle utilizzando gli editor SQL per Amazon Athena o Amazon Redshift. Per ulteriori informazioni, consulta Creazione di viste.	25 novembre 2023
Aggiornata la modifica alla politica	Ha documentato la modifica alla AWSLakeFormationCrossAccountManagerpolitica.	25 ottobre 2023

Documentazione aggiunta per la modalità di accesso ibrido	La modalità di accesso ibrido offre la flessibilità necessari a per abilitare selettivamente le autorizzazioni di Lake Formation per database e tabelle del tuo. AWS Glue Data Catalog Con la modalità di accesso ibrido, ora disponi di un percorso incrementale che ti consente di impostare le autorizzazioni di Lake Formation per un set specifico di utenti senza interromp ere le politiche di autorizza zione di altri utenti o carichi di lavoro esistenti. Per ulteriori informazioni, consulta Modalità di accesso ibrida.	26 settembre 2023
È stata aggiunta documenta zione per la creazione di tabelle Apache Iceberg	Ora puoi creare tabelle Apache Iceberg che utilizzan o il formato di dati Apache Parquet AWS Glue Data Catalog con dati che risiedono in Amazon S3. Per ulteriori informazioni, consulta Creazione di tabelle Iceberg.	16 agosto 2023

Documentazione aggiunta per l'accesso ai dati tra regioni	Lake Formation supporta l'interrogazione delle tabelle del Data Catalog tra AWS le regioni. Puoi accedere ai dati in una regione da altre regioni utilizzando Athena, Amazon EMR ed eseguire AWS Glue ETL creando collegamenti di risorse in altre regioni che puntano ai database e alle tabelle di origine. Puoi connettere il Data Catalog a metastore esterni che archiviano i metadati per i tuoi dati Amazon S3 e gestire in modo sicuro le autorizzazioni di accesso ai dati utilizzan do. AWS Lake Formation Per ulteriori informazioni, consulta Accesso alle tabelle tra regioni.	30 giugno 2023
Contenuti riorganizzati	Capitoli riorganizzati della guida per adattarli al percorso utente di Lake Formation.	15 maggio 2023

Documentazione aggiunta per la federazione HMS	Puoi connettere il Data Catalog a metastore esterni che archiviano i metadati per i tuoi dati Amazon S3 e gestire in modo sicuro le autorizzazioni di accesso ai dati utilizzando. AWS Lake Formation Per ulteriori informazioni, consulta Gestione delle autorizzazioni sui set di dati che utilizzano metastore esterni.	15 aprile 2023
Documentazione aggiunta per la condivisione dei dati di Amazon Redshift	Ora puoi gestire in modo sicuro i dati in un datashare di Amazon Redshift utilizzan do le autorizzazioni di Lake Formation. Lake Formation supporta la concessione di licenze di accesso ai tuoi dati tramiteAWS Data Exchange. Per ulteriori informazioni, consulta Condivisione dei dati in AWS Lake Formation.	30 novembre 2022
Support per la condivisione dei dati tra account direttamente con i mandanti	Sono state aggiunte informazi oni sulla condivisione diretta dei dati con i responsab ili IAM in un altro account. Per ulteriori informazioni, consulta Condivisione dei dati tra account in. AWS Lake Formation	10 novembre 2022

Support per la condivisione AWS RAM abilitata dei dati tramite TBAC	Sono state aggiunte informazioni sul metodo LF-TBAC per la concessione delle autorizzazioni di Data Catalog da utilizzare per le concessioni tra account. AWS Resource Access Manager	10 novembre 2022
È stata aggiunta una sezione sull'utilizzo di altri servizi	Sono state aggiunte informazioni su come AWS servizi come Athena, AWS Glue Redshift Spectrum e Amazon EMR possono utilizzare Lake Formation per accedere in modo sicuro ai dati nelle sedi Amazon S3 registrate con Lake Formation. Per ulteriori informazioni, consulta Collaborazione con altri AWS servizi.	10 novembre 2022
???	Sono state aggiunte informazioni sulla risoluzione di un errore durante l'utilizzo di Amazon EMR per accedere ai dati di più account. Per ulteriori informazioni, consulta Errore durante l'utilizzo di Amazon EMR per accedere ai dati condivisi tramite più account.	7 novembre 2022

Aggiornamenti alla condivisione di risorse tra account	È stata aggiunta una descrizione del funzionamento delle condivisioni di risorse tra account in Lake Formation. Ha documentato la modifica alla AWSLakeFormationCrossAccountManagerpolitica.	6 maggio 2022
Nuovi tutorial	Sono stati aggiunti nuovi tutorial per la creazione di tabelle gestite, la protezione e dei data lake e la condivisione dei data lake. Per ulteriori dettagli, consulta la sezione Guida introduttiva.	20 aprile 2022
Pagina iniziale di New Lake Formation	È stata aggiornata la landing page di Lake Formation per includere collegamenti a tutorial che forniscono step-by-step istruzioni su come creare un data lake, importare dati, condividere e proteggere i data lake utilizzando Lake Formation.	20 aprile 2022
Support per la vendita di credenziali	Sono state aggiunte informazioni sulla vendita di credenziali, che supporta Lake Formation per consentire ai servizi di terze parti di integrarsi con Lake Formation utilizzando le operazioni API di vendita delle credenziali. Per ulteriori informazioni, consulta Come funziona il vending di credenziali in Lake Formation.	28 febbraio 2022

Support per tabelle gestite e filtraggio avanzato dei dati	Sono state aggiunte informazioni sulle tabelle governate, che supportano transazioni ACID, compattazione automatica dei dati e query con viaggi nel tempo. Sono state aggiunte informazioni sulla creazione di filtri di dati per supportare la sicurezza a livello di colonna, la sicurezza a livello di riga e la sicurezza a livello di cella. Per ulteriori informazioni, consulta Governed Tables in Lake Formation and Data Filtering and Cell-Level Security in Lake Formation.	30 novembre 2021
Support per endpoint di interfaccia VPC	Sono state aggiunte informazioni sulla creazione di un endpoint di interfaccia cloud privato virtuale (VPC) per Lake Formation, in modo che la comunicazione tra il tuo VPC e Lake Formation avvenga in modo completo e sicuro all'interno della rete. AWS Per ulteriori informazioni, consulta Using Lake Formation with VPC Endpoints.	11 ottobre 2021

Supporto per le policy di endpoint VPC	Sono state aggiunte informazioni sul supporto per le policy degli endpoint Virtual Private Cloud (VPC) in Lake Formation. Per ulteriori informazioni, consulta Using Lake Formation with VPC Endpoints.	11 ottobre 2021
Support per il controllo degli accessi basato su tag	Il controllo degli accessi basato su tag Lake Formation offre un nuovo modo più scalabile per gestire l'accesso alle risorse del Data Catalog e ai dati sottostanti utilizzando i tag LF. Per ulteriori informazioni, consulta Lake Formation Tag-Based Access Control.	7 maggio 2021
Nuovo requisito di opt-in per il filtraggio dei dati su Amazon EMR.	Sono state aggiunte informazioni sul requisito di attivazione per consentire ad Amazon EMR di filtrare i dati gestiti da Lake Formation. Per ulteriori informazioni, consulta Consenti il filtraggio dei dati su Amazon EMR.	9 ottobre 2020
Support per la concessione di autorizzazioni complete per più account sui database Data Catalog	Sono state aggiunte informazioni sulla concessione delle autorizzazioni complete di Lake Formation sui database Data Catalog tra AWS gli account, tra cui. `CREATE_TABLE` Per ulteriori informazioni, consulta Sharing Data Catalog Databases.	1 ottobre 2020

Support per l'autenticazione Amazon Athena degli utenti tramite SAML.	Sono state aggiunte informazioni sul supporto per gli utenti Athena che si connettono tramite il driver JDBC o ODBC e si autenticano tramite provider di identità SAML come Okta e Microsoft Active Directory Federation Service (ADFS). Per ulteriori informazioni, consulta AWSService Integrations with Lake Formation.	30 settembre 2020
Support per l'accesso su più account con un catalogo dati crittografato	Sono state aggiunte informazioni sulla concessione di autorizzazioni per più account quando il Data Catalog è crittografato. Per ulteriori informazioni, consulta Prerequisiti per l'accesso su più account.	30 luglio 2020
Support per l'accesso al data lake da più account	Sono state aggiunte informazioni sulla concessione AWS Lake Formation delle autorizzazioni sui database e sulle tabelle di Data Catalog ad AWS account e organizzazioni esterni e sull'accesso agli oggetti Data Catalog condivisi da account esterni. Per ulteriori informazioni, vedere Cross-Account Access.	7 luglio 2020

Integrazione con Amazon QuickSight	Sono state aggiunte informazioni su come concedere le autorizzazioni di Lake Formation agli utenti di Amazon QuickSight Enterprise Edition in modo che possano accedere ai set di dati che risiedono in località Amazon S3 registrate. Per ulteriori informazioni, consulta Concessione delle autorizzazioni del catalogo dati.	29 giugno 2020
Aggiornamenti alla configurazione e ai capitoli introduttivi	Sono stati riorganizzati e migliorati i capitoli Configurazione e Guida introduttiva. Sono state aggiornate le autorizzazioni consigliate AWS Identity and Access Management (IAM) per l'amministratore del data lake.	27 febbraio 2020

Supporto perAWS Key Management Service	Sono state aggiunte informazioni su come il supporto di Lake Formation per AWS Key Management Service (AWS KMS) semplifica la configurazione di servizi integrati per leggere e scrivere dati crittografati nelle sedi registrate di Amazon Simple Storage Service (Amazon S3). Sono state aggiunte informazioni su come registrare posizioni Amazon S3 crittografate con. AWS KMS keys Per ulteriori informazioni, consulta the section called “Aggiungere una posizione Amazon S3 al tuo data lake”.	27 febbraio 2020
Aggiornamenti ai blueprint e alle politiche IAM degli amministratori del data lake	Parametri di input chiariti per i blueprint di database incrementali. Sono state aggiornate le politiche IAM richieste per un amministratore di data lake.	20 dicembre 2019
Riscrittura e aggiornamento del capitolo sulla sicurezza, revisioni del capitolo	Sono stati migliorati i capitoli relativi alla sicurezza e all'aggiornamento.	29 ottobre 2019
L'autorizzazione Super sostituisce tutte le autorizzazioni	Sono stati aggiornati i capitoli Sicurezza e Aggiornamento in modo da riflettere la sostituzione dell'autorizzazione con. `All` `Super`	10 ottobre 2019

Aggiunte, correzioni e chiarimenti	Sono state apportate aggiunte, correzioni e chiarimenti in base al feedback. È stato rivisto il capitolo sulla sicurezza. Sono stati aggiornati i capitoli Sicurezza e Aggiornamento in modo da rispecchiare la sostituzione del gruppo con. `Everyone` `IAMAllowedPrincipals`	11 settembre 2019
Nuova guida	Questa è la versione iniziale della Guida per gli sviluppatori di AWS Lake Formation.	8 agosto 2019

Glossario per AWS

Per la terminologia AWS più recente, consultare il glossario AWS nella documentazione di riferimento per Glossario AWS.

Le traduzioni sono generate tramite traduzione automatica. In caso di conflitto tra il contenuto di una traduzione e la versione originale in Inglese, quest'ultima prevarrà.