開發人員指南

AWS Lake Formation

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

AWS Lake Formation: 開發人員指南

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

Amazon 的商標和商業外觀不得用於任何非 Amazon 的產品或服務，也不能以任何可能造成客戶混淆、任何貶低或使 Amazon 名譽受損的方式使用 Amazon 的商標和商業外觀。所有其他非 Amazon 擁有的商標均為其各自擁有者的財產，這些擁有者可能附屬於 Amazon，或與 Amazon 有合作關係，亦或受到 Amazon 贊助。

Table of Contents

什麼是 AWS Lake Formation？ 1
Lake Formation 功能 1
資料擷取和管理 2
安全管理 3
將資料帶入 Data Catalog 4
運作方式 5
Lake Formation 許可管理工作流程 5
中繼資料許可 7
儲存存取管理 9
Lake Formation 中的跨帳戶資料共用 10
Lake Formation 元件 11
Lake Formation 主控台 11
Lake Formation API 和命令列界面 11
AWS 其他服務 11
Lake Formation 術語 12
資料湖 12
資料存取 12
混合存取模式 12
藍圖 12
工作流程 13
Data Catalog 13
基礎資料 13
Principal 13
資料湖管理員 13
AWS 服務與 Lake Formation 整合 14
其他 Lake Formation 資源 15
部落格 16
技術講座和網路研討會 16
現代架構 16
資料網格資源 16
最佳實務指南 16
Lake Formation 入門 17
開始使用 18
完成初始 AWS 組態任務 18

註冊 AWS 帳戶 18
建立具有管理存取權的使用者 19
授與程式設計存取權 20
設定 AWS Lake Formation 21
使用 AWS CloudFormation 範本設定 Lake Formation 資源 22
建立資料湖管理員 23
變更預設許可模型或使用混合存取模式 27
將許可指派給 Lake Formation 使用者 28
為您的資料湖設定 Amazon S3 位置 29
（選用）外部資料篩選設定 30
（選用）授予對 Data Catalog 加密金鑰的存取權 31
（選用）建立工作流程的 IAM 角色 31
將AWS Glue資料許可升級至 Lake Formation 模型 32
關於預設許可 33
列出現有的許可 34
設定 Lake Formation 許可 36
授予使用者 IAM 許可 37
切換到 Lake Formation 許可模型 37
步驟 5：保護新的 Data Catalog 資源 40
步驟 6：提供使用者新的 IAM 政策 41
步驟 7：清除現有的 IAM 政策 42
設定 Amazon VPC 端點 (AWS PrivateLink) 42
Lake Formation VPC 端點的考量事項 42
為 Lake Formation 建立介面 VPC 端點 42
為 Lake Formation 建立 VPC 端點政策 43
教學課程 45
從 AWS CloudTrail 來源建立資料湖 46
目標對象 47
先決條件 47
步驟 1：建立資料分析師使用者 48
步驟 2：將讀取 AWS CloudTrail 日誌的許可新增至工作流程角色 49
步驟 3：為資料湖建立 Amazon S3 儲存貯體 49
步驟 4：註冊 Amazon S3 路徑 50
步驟 5：授予資料位置許可 50
步驟 6：在 Data Catalog 中建立資料庫 50
步驟 7：授予資料許可 51

步驟 8：使用藍圖建立工作流程 51
步驟 9：執行工作流程 52
步驟 10：在資料表上授予 SELECT 53
步驟 11：使用 查詢資料湖 Amazon Athena 54
從 JDBC 來源建立資料湖 54
目標對象 55
先決條件 56
步驟 1：建立資料分析師使用者 56
步驟 2：在 中建立連線 AWS Glue 57
步驟 3：為資料湖建立 Amazon S3 儲存貯體 58
步驟 4：註冊 Amazon S3 路徑 58
步驟 5：授予資料位置許可 58
步驟 6：在 Data Catalog 中建立資料庫 59
步驟 7：授予資料許可 59
步驟 8：使用藍圖建立工作流程 59
步驟 9：執行工作流程 61
步驟 10：在資料表上授予 SELECT 62
步驟 11：使用 查詢資料湖 Amazon Athena 62
步驟 12：使用 Amazon Redshift Spectrum 查詢資料湖中的資料 63
步驟 13：使用 Amazon Redshift Spectrum 授予或撤銷 Lake Formation 許可 67
在 Lake Formation 中設定開放資料表格式的許可 67
目標對象 67
先決條件 68
步驟 1：佈建您的 資源 69
步驟 2：設定 Iceberg 資料表的許可 71
步驟 3：設定 Hudi 資料表的許可 77
步驟 4：設定 Delta Lake 資料表的許可 79
步驟 5：清除 AWS 資源 81
使用標籤型存取控制管理資料湖 82
目標對象 83
先決條件 84
步驟 1：佈建您的 資源 84
步驟 2：註冊您的資料位置、建立 LF 標籤本體，以及授予許可 85
步驟 3：建立 Lake Formation 資料庫 88
步驟 4：授予資料表許可 98
步驟 5：在 Amazon Athena 中執行查詢以驗證許可 100

步驟 6：清除 AWS 資源 101
使用資料列層級存取控制保護資料湖 101
目標對象 102
先決條件 102
步驟 1：佈建您的 資源 103
步驟 2：不含資料篩選條件的查詢 104
步驟 3：設定資料篩選條件並授予許可 106
步驟 4：使用資料篩選條件查詢 108
步驟 5：清除 AWS 資源 109
使用 Lake Formation 安全地共用您的資料 109
目標對象 110
設定 Lake Formation 設定 111
步驟 1：使用 AWS CloudFormation 範本佈建資源 113
步驟 2：Lake Formation 跨帳戶共用先決條件 115
步驟 3：使用標籤型存取控制方法實作跨帳戶共用 118
步驟 4：實作具名資源方法 123
步驟 5：清除 AWS 資源 126
AWS 帳戶 使用精細存取控制與外部 共用 Data Catalog 資源 127
目標對象 128
先決條件 129
步驟 1：提供對另一個 帳戶的精細存取 130
步驟 2：為相同帳戶中的使用者提供精細存取 131
加入 Lake Formation 許可 133
Lake Formation 許可概觀 134
精細存取控制的方法 135
中繼資料存取控制 137
基礎資料存取控制 140
Lake Formation 角色和 IAM 許可參考 145
AWS Lake Formation 角色 145
AWS Lake Formation 的 受管政策 146
角色建議的許可 154
變更資料湖的預設設定 164
隱含 Lake Formation 許可 167
Lake Formation 許可參考 169
每個資源類型的 Lake Formation 許可 169
Lake Formation 授予和撤銷 AWS CLI 命令 172

Lake Formation 許可 176
整合 IAM Identity Center 190
IAM Identity Center 與 Lake Formation 整合的先決條件 191
將 Lake Formation 與 IAM Identity Center 連線 194
更新 IAM Identity Center 整合 197
刪除 Lake Formation 與 IAM Identity Center 的連線 198
將許可授予使用者和群組 199
在 CloudTrail 日誌中包含 IAM Identity Center 使用者內容 202
將 Amazon S3 位置新增至您的資料湖 204
用於註冊位置的角色需求 205
註冊 Amazon S3 位置 211
註冊加密的 Amazon S3 位置 214
在另一個 AWS 帳戶中註冊 Amazon S3 位置 218
跨 AWS 帳戶註冊加密的 Amazon S3 位置 220
取消註冊 Amazon S3 位置 224
混合存取模式 225
常見的混合存取模式使用案例 226
混合存取模式的運作方式 227
設定混合存取模式 - 常見案例 229
從混合存取模式移除主體和資源 246
在混合存取模式中檢視主體和資源 247
其他資源 248
在 中建立物件 AWS Glue Data Catalog 248
建立目錄 249
建立資料庫 250
建立資料表 250
建置資料目錄檢視 257
使用工作流程匯入資料 288
藍圖和工作流程 289
建立工作流程 290
執行工作流程 293
將資料帶入 Data Catalog 295
將 Amazon Redshift 資料帶入 Data Catalog 296
主要優點 298
角色和責任 298
先決條件 299

建立 Amazon Redshift 聯合目錄 305
檢視目錄物件 313
更新聯合目錄 315
存取共用聯合目錄 316
刪除聯合目錄 320
查詢聯合目錄 321
其他資源 322
聯合到外部資料來源 322
工作流程 323
先決條件 323
建立聯合目錄 327
檢視目錄物件 332
刪除聯合目錄 334
查詢聯合目錄 335
其他資源 335
在 Data Catalog 中建立 Amazon S3 資料表目錄 336
Data Catalog 和 Lake Formation 整合的運作方式 336
先決條件 337
啟用 Amazon S3 Tables 整合 339
建立資料庫和資料表 342
在另一個 AWS 帳戶中註冊 Amazon S3 資料表儲存貯體 345
授予許可 348
建立 Amazon Redshift 受管目錄 350
管理 Amazon Redshift 資料共用中資料的許可 354
先決條件 355
設定 Amazon Redshift 資料共用的許可 356
查詢聯合資料庫 360
管理使用外部中繼存放區的資料集許可 360
工作流程 362
先決條件 363
將 Data Catalog 連接到外部 Hive 中繼存放區 366
其他資源 369
管理 Lake Formation 許可 370
授予資料位置許可 370
授予資料位置許可 （相同帳戶） 371
授予資料位置許可 （外部帳戶） 373

授予與您帳戶共用之資料位置的許可 376
授予資料許可 377
授予 Lake Formation 許可所需的 IAM 許可 378
使用具名資源方法 380
標籤式存取控制 403
使用 LF-TBAC 方法授予資料湖許可 459
屬性型存取控制 466
許可範例案例 469
資料篩選和儲存格層級安全性 471
資料篩選條件 472
資料列篩選條件表達式中的 PartiQL 支援 476
使用儲存格層級篩選查詢資料表所需的許可 478
管理資料篩選條件 479
檢視資料庫和資料表許可 494
使用主控台撤銷許可 497
跨帳戶資料共用 498
先決條件 500
更新跨帳戶資料共用版本設定 504
從外部帳戶跨 AWS 帳戶 或 IAM 主體共用 Data Catalog 資料表和資料庫 509
授予與您帳戶共用之資料庫或資料表的許可 514
授予資源連結許可 516
存取共用資料表的基礎資料 518
跨帳戶 CloudTrail 記錄 520
使用 AWS Glue和 Lake Formation 管理跨帳戶許可 524
使用 GetResourceShares API 操作檢視所有跨帳戶授予 527
存取和檢視共用的 Data Catalog 資料表和資料庫 528
接受 AWS RAM 資源共享邀請 529
檢視共用的 Data Catalog 資料表和資料庫 531
建立資源連結 532
資源連結的運作方式 533
建立共用資料表的資源連結 535
建立共用資料庫的資源連結 538
AWS Glue APIs中的資源連結處理 541
跨區域存取資料表 544
工作流程 545
設定跨區域資料表存取 549

安全 552
資料保護 552
靜態加密 553
基礎設施安全性 554
預防跨服務混淆代理人 554
安全事件登入 AWS Lake Formation 555
與 Lake Formation 整合 556
使用 Lake Formation 應用程式整合 556
Lake Formation 應用程式整合的運作方式 557
Lake Formation 應用程式整合中的角色和責任 558
Lake Formation 應用程式整合 API 操作的工作流程 559
註冊第三方查詢引擎 560
啟用第三方查詢引擎呼叫應用程式整合 API 操作的許可 561
完整資料表存取的應用程式整合 566
使用其他 AWS 服務 569
Amazon Athena 572
支援交易資料表格式 573
其他資源 575
Amazon Redshift Spectrum 575
支援交易資料表類型 576
其他資源 577
AWS Glue 577
支援交易資料表類型 578
其他資源 579
Amazon EMR 579
支援交易資料表格式 580
其他資源 581
QuickSight 581
其他資源 582
AWS CloudTrail 湖 582
使用 記錄 AWS Lake Formation API 呼叫 AWS CloudTrail 583
CloudTrail 中的 Lake Formation 資訊 583
了解 Lake Formation 事件 584
Lake Formation 最佳實務、考量和限制 586
跨帳戶資料共用最佳實務和考量事項 586
服務連結角色限制 588

跨區域資料存取限制 588
Data Catalog 檢視的考量和限制 589
資料篩選限制 590
資料欄層級篩選的備註和限制 590
儲存格層級篩選限制 591
混合存取模式的考量和限制 593
將 Amazon Redshift 資料倉儲資料帶入 的限制 AWS Glue Data Catalog 594
S3 資料表目錄整合限制 595
Hive 中繼資料存放區資料共用的考量和限制 596
Amazon Redshift 資料共用限制 597
IAM Identity Center 整合限制 598
Lake Formation 標籤型存取控制最佳實務和考量事項 598
屬性型存取控制考量事項、限制和支援的區域 601
Lake Formation 故障診斷 602
一般性問題的故障診斷 602
錯誤：<Amazon S3 location> 上的 Lake Formation 許可不足 602
錯誤：「Glue API 的加密金鑰許可不足」 602
使用資訊清單的我 Amazon Athena 或 Amazon Redshift 查詢失敗 602
錯誤：「Lake Formation (Lake Formation) 許可不足：目錄上的必要建立標籤」 603
刪除無效的資料湖管理員時發生錯誤 603
對跨帳戶存取進行故障診斷 603
我授予跨帳戶 Lake Formation 許可，但收件人看不到資源 603
收件人帳戶中的主體可以查看 Data Catalog 資源，但無法存取基礎資料 604
錯誤：接受 AWS RAM 資源共享邀請時，「關聯失敗，因為發起人未授權」 604
錯誤：「未授權授予資源的許可」 605
錯誤：「拒絕擷取 AWS 組織資訊的存取」 605
錯誤：「找不到組織 <organization-ID>」 605
錯誤：「Lake Formation 許可不足：組合無效」 605
對外部帳戶授予/撤銷請求的 ConcurrentModificationException 605
使用 Amazon EMR 存取跨帳戶共用的資料時發生錯誤 605
疑難排解藍圖和工作流程 606
我的藍圖失敗，「使用者：<user-ARN> 未獲授權執行：iam：PassRole on resource：<role-ARN>」 607
我的工作流程失敗，「使用者：<user-ARN> 未獲授權執行：iam：PassRole on resource：<role-ARN>」 607
我工作流程中的爬蟲程式失敗，出現「資源不存在或請求者無權存取請求的許可」 607

工作流程中的爬蟲程式失敗，並顯示「呼叫 CreateTable 操作時發生錯誤 (AccessDeniedException)...」 607
的已知問題 AWS Lake Formation 607
篩選資料表中繼資料的限制 608
重新命名排除資料欄時的問題 609
刪除 CSV 資料表中資料欄的問題 609
資料表分割區必須新增至通用路徑下 609
在工作流程建立期間建立資料庫的問題 609
刪除並重新建立使用者的問題 609
Data Catalog API 操作不會更新 IsRegisteredWithLakeFormation 參數的值 610
Lake Formation 操作不支援 AWS Glue 結構描述登錄檔 610
已更新錯誤訊息 610
Lake Formation API 611
許可 612
— operations — 612
— 資料類型 — 612
資料湖設定 613
— operations — 613
— 資料類型 — 613
IAM Identity Center 整合 613
— operations — 613
— 資料類型 — 613
混合存取模式 613
— operations — 613
— 資料類型 — 612
登入資料販賣 614
— operations — 614
— 資料類型 — 614
標記 615
— operations — 615
— 資料類型 — 615
資料篩選條件 APIs 616
— operations — 616
— 資料類型 — 616
常見資料類型 616
ErrorDetail 616

字串模式 617
支援地區 618
一般可用性 618
AWS GovCloud (US) 618
交易和儲存最佳化 618
文件歷史記錄 621
AWS 詞彙表 631
..... dcxxxii

什麼是 AWS Lake Formation？

歡迎使用 AWS Lake Formation 開發人員指南。

AWS Lake Formation 可協助您集中管理、保護和全域共用資料，以進行分析和機器學習。使用 Lake Formation，您可以管理 Amazon Simple Storage Service (Amazon S3) 及其中繼資料上資料湖資料的精細存取控制 AWS Glue Data Catalog。

Lake Formation 提供自己的許可模型，可增強 IAM 許可模型。Lake Formation 許可模型可透過簡單的授予或撤銷機制，對存放在資料湖中的資料以及外部資料來源進行精細存取，例如 Amazon Redshift 資料倉儲、 Amazon DynamoDB 資料庫和第三方資料來源，就像關聯式資料庫管理系統 (RDBMS)。Lake Formation 許可是使用在 AWS 分析和機器學習服務的資料欄、資料列和儲存格層級的精細控制項強制執行，包括 Amazon Athena Amazon QuickSight、Amazon Redshift Spectrum、Amazon EMR 和 AWS Glue。

使用適用於 AWS Glue Data Catalog (Data Catalog) 的 Lake Formation 混合存取模式，您可以使用 Lake Formation 許可和 Amazon S3 和 動作的 IAM 許可政策來保護和 AWS Glue 存取目錄資料。透過混合存取模式，資料管理員可以選擇性地和遞增地加入 Lake Formation 許可，一次專注於一個資料湖使用案例。

Lake Formation 也可讓您在內部與外部跨多個 AWS 帳戶、 AWS 組織共用資料，或直接與另一個帳戶中的 IAM 主體共用資料，提供對 Data Catalog 中繼資料和基礎資料的精細存取。

主題

- Lake Formation 功能
- AWS Lake Formation：運作方式
- Lake Formation 元件
- Lake Formation 術語
- AWS 服務與 Lake Formation 整合
- 其他 Lake Formation 資源
- Lake Formation 入門

Lake Formation 功能

Lake Formation 可協助您細分資料孤島，並將不同類型的結構化和非結構化資料合併到集中式儲存庫中。首先，識別 Amazon S3 或關聯式和 NoSQL 資料庫中的現有資料存放區，並將資料移至您的資料

湖。然後編目、編目和準備資料以供分析。接下來，透過使用者選擇的分析服務，為您的使用者提供對資料的安全自助式存取。

您可以使用 Lake Formation 主控台在 Data Catalog 中建立多層聯合目錄，並在 Amazon S3 資料湖和 Amazon Redshift 資料倉儲中統一資料。您也可以整合操作資料庫中的資料 Amazon DynamoDB，例如 Google BigQuery、MySQL 等第三方資料來源。Data Catalog 提供集中式中繼資料儲存庫，可讓您更輕鬆地管理和探索不同系統中的資料。

如需詳細資訊，請參閱將您的資料帶入 AWS Glue Data Catalog。

主題

- 資料擷取和管理
- 安全管理
- 將資料帶入 Data Catalog

資料擷取和管理

從 中已存在的資料庫匯入資料 AWS

指定現有資料庫的位置並提供存取登入資料後，Lake Formation 會讀取資料及其中繼資料 （結構描述），以了解資料來源的內容。然後，它會將資料匯入新的資料湖，並將中繼資料記錄在中央目錄中。使用 Lake Formation，您可以從在 Amazon RDS 中執行或在 Amazon EC2 中託管的 MySQL、PostgreSQL、SQL Server、MariaDB 和 Oracle 資料庫匯入資料。支援大量和增量資料載入。

從其他外部來源匯入資料

您可以使用 Lake Formation 與 Java Database Connectivity (JDBC) 連線，從內部部署資料庫移動資料。識別您的目標來源，並在主控台中提供存取憑證，Lake Formation 會讀取您的資料並將其載入資料湖。若要從上述資料庫以外的資料庫匯入資料，您可以使用 建立自訂 ETL 任務 AWS Glue。

編目和標記您的資料

您可以使用 AWS Glue 爬蟲程式讀取 Amazon S3 中的資料，並擷取資料庫和資料表結構描述，並將該資料存放在可搜尋的資料目錄中。然後，使用 Lake Formation Lake Formation 標籤型存取控制(TBAC) 來管理資料庫、資料表和資料欄的許可。如需將資料表新增至 Data Catalog 的詳細資訊，請參閱 在 中建立物件 AWS Glue Data Catalog。

安全管理

定義和管理存取控制

Lake Formation 提供單一位置來管理資料湖中資料的存取控制。您可以定義安全政策，限制對資料庫、資料表、資料欄、資料列和儲存格層級資料的存取。這些政策適用於 IAM 使用者和角色，以及透過外部身分提供者聯合時的使用者和群組。您可以使用精細控制來存取 Amazon Redshift Spectrum、Athena、 AWS Glue ETL 和 Amazon EMR for Apache Spark 中 Lake Formation 保護的資料。每當您建立 IAM 身分時，請務必遵循 IAM 最佳實務。如需詳細資訊，請參閱《IAM 使用者指南》中的安全最佳實務。

混合存取模式

Lake Formation 混合存取模式提供彈性，讓您選擇性地為 Data Catalog 中的資料庫和資料表啟用 Lake Formation 許可。使用混合存取模式時，您現在有一個增量路徑，可讓您為一組特定使用者設定 Lake Formation 許可，而不會中斷其他現有使用者或工作負載的許可政策。如需詳細資訊，請參閱混合存取模式。

實作稽核記錄

Lake Formation 透過 CloudTrail 提供全面的稽核日誌，以監控存取並顯示是否符合集中定義的政策。您可以跨分析和機器學習服務稽核資料存取歷史記錄，這些服務會透過 Lake Formation 讀取資料湖中的資料。這可讓您查看哪些使用者或角色嘗試存取哪些資料、使用哪些 服務以及何時存取。您可以使用與使用 CloudTrail APIs 和主控台存取任何其他 CloudTrail 日誌相同的方式來存取稽核日誌。如需 CloudTrail 日誌的詳細資訊，請參閱 使用 記錄 AWS Lake Formation API 呼叫 AWS CloudTrail。

資料列和儲存格層級安全性

Lake Formation 提供資料篩選條件，可讓您限制對資料欄和資料列組合的存取。使用資料列和儲存格層級安全性來保護敏感資料，例如個人身分識別資訊 (PII)。如需資料列層級安全性的詳細資訊，請參閱 Lake Formation 中的資料篩選和儲存格層級安全性。

標籤式存取控制

使用 Lake Formation 標籤型存取控制，透過建立稱為 LF 標籤的自訂標籤來管理數百甚至數千個資料許可。您現在可以定義 LF 標籤，並將其連接到資料庫、資料表或資料欄。然後，跨分析、機器學習 (ML) 和擷取、轉換和載入 (ETL) 服務共用受控存取以供取用。LF-Tags 使用幾個邏輯標籤取代數千個資源的政策定義，以確保可以輕鬆擴展資料控管。Lake Formation 對此中繼資料提供文字型搜尋，因此您的使用者可以快速找到他們需要分析的資料。

跨帳戶存取

Lake Formation 許可管理功能透過集中式方法簡化跨多個 AWS 帳戶的分散式資料湖的安全和管理，為 Data Catalog 和 Amazon S3 位置提供精細的存取控制。如需詳細資訊，請參閱Lake Formation 中的跨帳戶資料共用。

將資料帶入 Data Catalog

聯合功能可讓您建立聯合型錄，並設定存放在 Amazon Redshift 等不同資料來源中的資料集的許可，而無需將資料或中繼資料遷移至 Amazon S3 或 AWS Glue Data Catalog。您可以使用下列方法，在 Lake Formation 中為外部資料集帶來資料和管理許可：

如需詳細資訊，請參閱將資料帶入 AWS Glue Data Catalog。

- 將 Amazon Redshift 資料倉儲中的資料帶入 AWS Glue Data Catalog：向 Data Catalog 註冊現有的 Amazon Redshift 命名空間或叢集，並在 Data Catalog 中建立多層聯合目錄。

  您可以使用任何與 Apache Iceberg REST 目錄 OpenAPI 規格相容的查詢引擎來存取您的資料，例如 Amazon EMR Serverless 和 Amazon Athena。

  如需詳細資訊，請參閱將 Amazon Redshift 資料帶入 AWS Glue Data Catalog。
- 從外部資料來源聯合到 Data Catalog – 使用連線將 Data Catalog 連接到外部資料來源 AWS Glue ，並使用 Lake Formation 建立聯合目錄以集中管理資料集的存取許可。無需將中繼資料遷移至 Data Catalog。

  如需詳細資訊，請參閱在 中聯合到外部資料來源 AWS Glue Data Catalog。
- 將 Amazon S3 資料表儲存貯體與 Data Catalog 整合 – 您可以將 Amazon S3 資料表發佈並編目為 Data Catalog 物件，並從 Lake Formation 主控台或使用 AWS Glue APIs將目錄註冊為 Lake Formation 資料位置。

  如需詳細資訊，請參閱在 中建立 Amazon S3 Tables 目錄 AWS Glue Data Catalog。
- 建立目錄以管理 Data Catalog 中的 Amazon Redshift 資料表 – 您目前可能沒有可用的 Amazon Redshift 生產者叢集或 Amazon Redshift 資料共用，但想要使用 Data Catalog 建立和管理 Amazon Redshift 資料表。您可以使用 `glue:CreateCatalog` API 或 AWS Lake Formation 主控台建立 AWS Glue 受管目錄，將目錄類型設定為 `Managed`和 `Catalog source` Redshift，以開始使用。

  如需詳細資訊，請參閱在 中建立 Amazon Redshift 受管目錄 AWS Glue Data Catalog。
- 將 Lake Formation 與 Amazon Redshift 資料共用整合 – 使用 Lake Formation 集中管理 Amazon Redshift 資料共用的資料庫、資料表、資料欄和資料列層級存取許可，並限制使用者存取資料共用中的物件。

- 將 Data Catalog 連線至外部中繼存放區 – AWS Glue Data Catalog 連線至外部中繼存放區，以使用 Lake Formation 管理 Amazon S3 中資料集的存取許可。無需將中繼資料遷移至 Data Catalog。

  如需詳細資訊，請參閱管理使用外部中繼存放區的資料集許可。

- 將 Lake Formation 與 AWS 資料交換整合 – Lake Formation 支援透過 授權存取您的資料 AWS Data Exchange。如果您有興趣授權 Lake Formation 資料，請參閱AWS Data Exchange 《 使用者指南》中的什麼是 AWS Data Exchange 。

AWS Lake Formation：運作方式

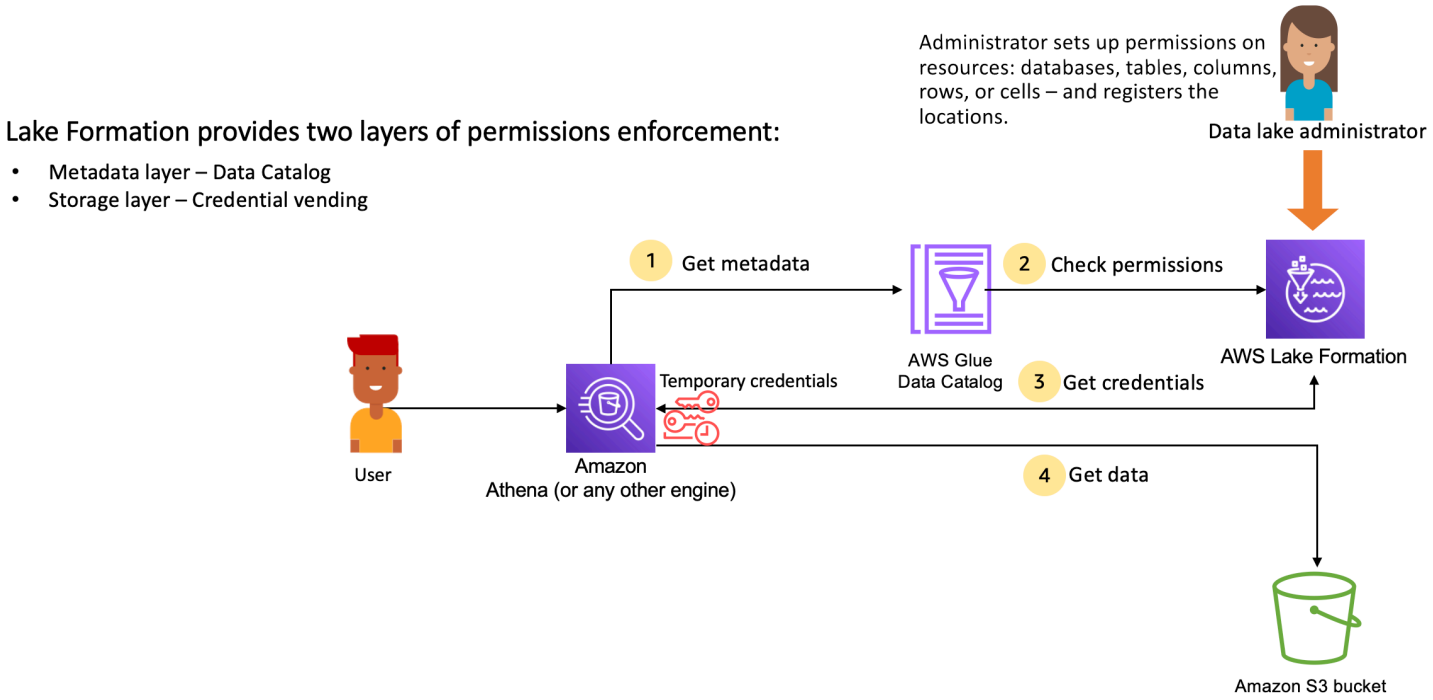
AWS Lake Formation 提供關聯式資料庫管理系統 (RDBMS) 許可模型，以授予或撤銷對 Data Catalog 資源的存取權，例如 Amazon S3 中具有基礎資料的資料庫、資料表和資料欄。易於管理的 Lake Formation 許可會取代複雜的 Amazon S3 儲存貯體政策和對應的 IAM 政策。

在 Lake Formation 中，您可以在兩個層級實作許可：

- 對資料庫和資料表等 Data Catalog 資源強制執行中繼資料層級許可
- 代表整合引擎管理存放在 Amazon S3 中基礎資料的儲存存取許可

Lake Formation 許可管理工作流程

Lake Formation 與分析引擎整合，以查詢向 Lake Formation 註冊的 Amazon S3 資料存放區和中繼資料物件。下圖說明許可管理如何在 Lake Formation 中運作。



Lake Formation 許可管理高階步驟

在 Lake Formation 可以為資料湖中的資料提供存取控制之前，具有管理許可的資料湖管理員或使用者會設定個別 Data Catalog 資料表使用者政策，以允許或拒絕使用 Lake Formation 許可存取 Data Catalog 資料表。

然後，資料湖管理員或管理員委派的使用者會將 Lake Formation 許可授予 Data Catalog 資料庫和資料表上的使用者，並向 Lake Formation 註冊資料表的 Amazon S3 位置。

1. 取得中繼資料 – 委託人（使用者）將查詢或 ETL 指令碼提交至整合的分析引擎，例如 Amazon Athena AWS Glue、Amazon EMR 或 Amazon Redshift Spectrum。整合式分析引擎會識別正在請求的資料表，並將中繼資料請求傳送至 Data Catalog。
2. 檢查許可 – Data Catalog 會使用 Lake Formation 檢查使用者的許可，如果使用者獲得存取資料表的授權， 會將允許使用者查看的中繼資料傳回給引擎。
3. 取得憑證 – Data Catalog 可讓引擎知道資料表是否由 Lake Formation 管理。如果基礎資料已向 Lake Formation 註冊，分析引擎會請求 Lake Formation 透過授予暫時存取權來提供資料存取。
4. 取得資料 – 如果使用者獲得存取資料表的授權，Lake Formation 會提供整合式分析引擎的暫時存取權。使用暫時存取，分析引擎會從 Amazon S3 擷取資料，並執行必要的篩選，例如資料欄、資料列或儲存格篩選。當引擎完成執行任務時，會將結果傳回給使用者。此程序稱為登入資料販賣。

   如果資料表不是由 Lake Formation 管理，則分析引擎的第二個呼叫會直接對 Amazon S3 進行。評估相關的 Amazon S3 儲存貯體政策和 IAM 使用者政策以進行資料存取。

每當您使用 IAM 政策時，請務必遵循 IAM 最佳實務。如需詳細資訊，請參閱《IAM 使用者指南》中的 IAM 中的安全性最佳實務。

主題

- 中繼資料許可
- 儲存存取管理
- Lake Formation 中的跨帳戶資料共用

中繼資料許可

Lake Formation 提供 Data Catalog 的授權和存取控制。當 IAM 角色從任何系統進行 Data Catalog API 呼叫時，Data Catalog 會驗證使用者的資料許可，並僅傳回使用者有權存取的中繼資料。例如，如果 IAM 角色只能存取資料庫中的一個資料表，且擔任該角色的服務或使用者執行`GetTables`操作，則無論資料庫中的資料表數目為何，回應只會包含一個資料表。

預設設定 - **`IAMAllowedPrincipal`**群組許可

AWS Lake Formation根據預設， 會將所有資料庫和資料表的許可設定為名為 的虛擬群組`IAMAllowedPrincipal`。此群組是唯一的，並且僅在 Lake Formation 中可見。`IAMAllowedPrincipal` 群組包含所有可透過 IAM 主體政策和資源政策存取 Data Catalog AWS Glue 資源的 IAM 主體。如果此許可存在於資料庫或資料表上，則會授予所有主體存取資料庫或資料表的權限。

如果您想要在資料庫或資料表上提供更精細的許可，請移除`IAMAllowedPrincipal`許可，Lake Formation 會強制執行與該資料庫或資料表相關聯的所有其他政策。例如，如果有政策允許使用者 A 存取具有`DESCRIBE`許可的資料庫 A，且 `IAMAllowedPrincipal` 具有所有許可，則使用者 A 將繼續執行所有其他動作，直到撤銷`IAMAllowedPrincipal`許可為止。

此外，根據預設， `IAMAllowedPrincipal`群組在建立所有新資料庫和資料表時都有其許可。控制此行為的組態有兩種。第一個位於帳戶和區域層級，為新建立的資料庫啟用此功能，第二個位於資料庫層級。若要修改預設設定，請參閱 變更預設許可模型或使用混合存取模式。

授予許可

資料湖管理員可以將 Data Catalog 許可授予主體，以便主體可以建立和管理資料庫和資料表，並可以存取基礎資料。

資料庫和資料表層級許可

當您在 Lake Formation 中授予許可時，授予者必須指定授予許可的委託人、授予許可的資源，以及承授者應有權執行的動作。對於 Lake Formation 中的大多數資源，授予許可的委託人清單和資源都很類似，但承授者可以執行的動作會根據資源類型而有所不同。例如，資料表可讀取資料表的SELECT許可，但資料庫不允許SELECT許可。資料庫上允許 CREATE_TABLE許可，但資料表上不允許 許可。

您可以使用兩種方法來授予 AWS Lake Formation 許可：

- 具名資源方法 – 可讓您在授予使用者許可時選擇資料庫和資料表名稱。
- LF 標籤型存取控制 (LF-TBAC) – 使用者建立 LF 標籤、將其與 Data Catalog 資源建立關聯、對 LF 標籤授予Describe許可、將許可關聯至個別使用者，以及使用 LF 標籤將 LF 許可政策寫入至不同使用者。這類以 LF-Tag-based政策適用於與這些 LF 標籤值相關聯的所有資料目錄資源。

> **Note**
>
> LF 標籤對 Lake Formation 是唯一的。它們僅在 Lake Formation 中可見，不應與 AWS 資源標籤混淆。

LF-TBAC 是一項功能，可讓使用者將資源分組為使用者定義的 LF 標籤類別，並在這些資源群組上套用許可。因此，這是跨大量 Data Catalog 資源擴展許可的最佳方式。

如需詳細資訊，請參閱Lake Formation 標籤型存取控制。

當您將許可授予委託人時，Lake Formation 會將許可評估為該使用者所有政策的聯集。例如，如果您的委託人資料表上有兩個政策，其中一個政策透過具名資源方法授予資料欄 col1、col2 和 col3 的許可，而另一個政策授予相同資料表和委託人 col5 的許可，以及透過 LF 標籤的 col6，則有效的許可將是許可的聯集，即 col1、col2、col3、col5 和 col6。這也包含資料篩選條件和資料列。

資料位置許可

資料位置許可可讓非管理使用者在特定 Amazon S3 位置建立資料庫和資料表。如果使用者嘗試在他們沒有建立許可的位置建立資料庫或資料表，則建立任務會失敗。這是為了防止使用者在資料湖內的任意位置建立資料表，並提供使用者可讀取和寫入資料的位置控制。在建立資料表的資料庫中的 Amazon S3 位置中建立資料表時，會有隱含許可。如需詳細資訊，請參閱授予資料位置許可。

建立資料表和資料庫許可

根據預設，非管理使用者沒有在資料庫中建立資料庫或資料表的許可。資料庫建立是在帳戶層級使用 Lake Formation 設定進行控制，因此只有授權的主體才能建立資料庫。如需詳細資訊，請參閱建立資

料庫。若要建立資料表，主體需要建立資料表之資料庫的CREATE_TABLE許可。如需詳細資訊，請參閱建立資料表。

隱含和明確許可

Lake Formation 會根據角色和角色執行的動作提供隱含許可。例如，資料湖管理員會自動取得資料目錄中所有資源的DESCRIBE許可、所有位置的資料位置許可、在所有位置建立資料庫和資料表的許可，Grant以及任何資源的Revoke許可。資料庫建立者會自動取得其所建立資料庫的所有資料庫許可，而資料表建立者則取得其所建立資料表的所有許可。如需詳細資訊，請參閱隱含 Lake Formation 許可。

准許許可

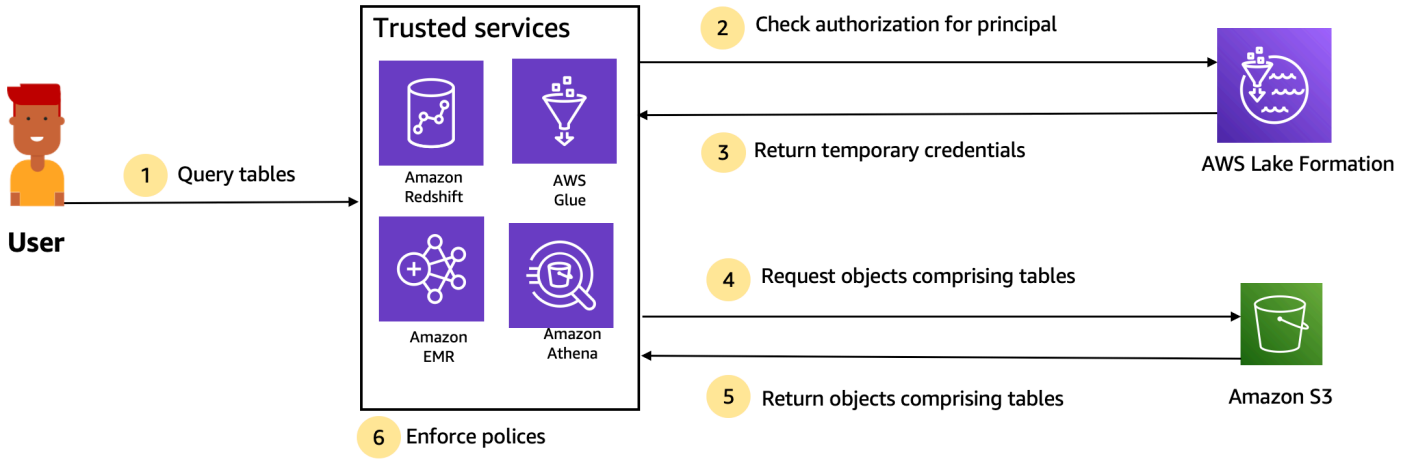
資料湖管理員能夠透過提供可授予的許可，將許可的管理委派給非管理使用者。當委託人獲得資源的可授予許可和一組許可時，該委託人將能夠授予該資源上的其他委託人許可。

儲存存取管理

Lake Formation 使用登入資料販賣功能來暫時存取 Amazon S3 資料。登入資料販賣或字符販賣是一種常見的模式，可提供臨時登入資料給使用者、服務或一些其他實體，以授予短期存取資源。

Lake Formation 會利用此模式，提供 Athena 等 AWS 分析服務的短期存取權，以代表呼叫委託人存取資料。授予許可時，使用者不需要更新其 Amazon S3 儲存貯體政策或 IAM 政策，也不需要直接存取 Amazon S3。

下圖顯示 Lake Formation 如何暫時存取已註冊的位置：



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

1. 委託人（使用者）透過 Athena、Amazon EMR、Redshift Spectrum 或 等受信任的整合服務輸入資料表的查詢或資料請求 AWS Glue。
2. 整合的服務會檢查 Lake Formation 的授權，以取得資料表和請求的資料欄，並進行授權判斷。如果使用者未獲授權，Lake Formation 會拒絕存取資料，且查詢會失敗。
3. 資料表和使用者的授權成功且開啟儲存授權後，整合服務會從 Lake Formation 擷取臨時憑證以存取資料。
4. 整合的服務會使用 Lake Formation 的臨時登入資料，從 Amazon S3 請求物件。
5. Amazon S3 提供整合服務的 Amazon S3 物件。Amazon S3 物件包含資料表中的所有資料。
6. 整合的服務會執行 Lake Formation 政策的必要強制執行，例如資料欄層級、資料列層級和/或儲存格層級篩選。整合的服務會處理查詢，並將結果傳回給使用者。

啟用 Data Catalog 資料表的儲存層級許可強制執行

根據預設，不會為 Data Catalog 中的資料表啟用儲存層級強制執行。若要啟用儲存層級強制執行，您必須向 Lake Formation 註冊來源資料的 Amazon S3 位置，並提供 IAM 角色。將針對具有相同資料表位置路徑或 Amazon S3 位置字首的所有資料表啟用儲存層級許可。

當整合的服務代表使用者請求存取資料位置時，Lake Formation 服務會擔任此角色，並將登入資料傳回至具有資源縮小範圍許可的請求服務，以便進行資料存取。已註冊的 IAM 角色必須具備 Amazon S3 位置的所有必要存取權，包括 AWS KMS 金鑰。

如需詳細資訊，請參閱註冊 Amazon S3 位置。

支援 AWS 的服務

AWS 分析服務，例如 Athena、Redshift Spectrum、Amazon EMR AWS Glue Amazon QuickSight，並使用 AWS Lake Formation 憑證販賣 API 操作與 Lake Formation Amazon SageMaker AI 整合。若要查看與 Lake Formation 整合 AWS 的服務完整清單，以及其支援的精細程度和資料表格式，請參閱使用其他 AWS 服務。

Lake Formation 中的跨帳戶資料共用

使用 Lake Formation，您可以使用具名資源方法或 LF 標籤，以簡單的設定在 AWS 帳戶內和跨帳戶共用 Data Catalog 資源（資料庫和資料表）。您可以將整個資料庫或從資料庫中選取資料表，分享給帳戶中的任何 IAM 主體 (IAM 角色和使用者）、帳戶層級的其他 AWS 帳戶，或直接分享給另一個帳戶中的 IAM 主體。

您也可以與資料篩選條件共用 Data Catalog 資料表，以限制對資料列層級和儲存格層級詳細資訊的存取。Lake Formation 使用 AWS Resource Access Manager (AWS RAM) 來協助在帳戶之間授予許

可。在兩個帳戶之間共用資源時， AWS RAM 會將邀請傳送給收件人帳戶。當使用者接受 AWS RAM 共享邀請時， AWS RAM 會提供 Lake Formation 必要的許可，讓 Data Catalog 資源可用，以及啟用的儲存層級強制執行。如需詳細資訊，請參閱Lake Formation 中的跨帳戶資料共用。

當收件人帳戶的資料湖管理員接受 AWS RAM 共用時，收件人帳戶中會提供共用資源。如果管理員擁有共用資源的許可，則資料湖管理員會將共用資源GRANTABLE的 Lake Formation 許可進一步授予收件人帳戶中的其他 IAM 主體。

不過，在沒有資源連結的情況下，主體無法使用 Athena 或 Redshift Spectrum 查詢共用資源。資源連結是 Data Catalog 中的實體，類似於 Linux-Symlink 概念。

收件人帳戶的資料湖管理員會在共用資源上建立資源連結。管理員授予 資源連結的Describe許可，以及原始共用資源所需的許可給其他使用者。然後，收件人帳戶中的使用者可以使用資源連結，使用 Athena 和 Redshift Spectrum 查詢共用資源。如需資源連結的詳細資訊，請參閱 建立資源連結。

Lake Formation 元件

AWS Lake Formation 依賴多個元件的互動來建立和管理資料湖。

Lake Formation 主控台

您可以使用 Lake Formation 主控台來定義和管理資料湖，並授予和撤銷 Lake Formation 許可。您可以在 主控台上使用藍圖來探索、清理、轉換和擷取資料。您也可以為個別 Lake Formation 使用者啟用或停用對 主控台的存取。

Lake Formation API 和命令列界面

Lake Formation 透過多種語言特定的 SDKs和 AWS Command Line Interface () 提供 API 操作AWS CLI。Lake Formation API 可與 AWS Glue API 搭配使用。Lake Formation API 主要著重於管理 Lake Formation 許可，而 AWS Glue API 則提供資料目錄 API 和受管基礎設施，用於定義、排程和執行資料上的 ETL 操作。

如需 AWS Glue API 的相關資訊，請參閱 AWS Glue 開發人員指南。如需使用 的詳細資訊 AWS CLI，請參閱 AWS CLI 命令參考。

AWS 其他服務

Lake Formation 使用以下服務：

- AWS Glue 協調任務和爬蟲程式，以使用轉換AWS Glue轉換資料。

- IAM 將許可政策授予 Lake Formation 主體。Lake Formation 許可模型增強了 IAM 許可模型，以保護您的資料湖。

Lake Formation 術語

以下是您將在本指南中遇到的一些重要術語。

資料湖

資料湖是存放在 Amazon S3 中的持久性資料，並由 Lake Formation 使用 Data Catalog 管理。資料湖通常會存放下列項目：

- 結構化和非結構化資料
- 原始資料和轉換後的資料

若要讓 Amazon S3 路徑位於資料湖內，則必須向 Lake Formation 註冊。

資料存取

Lake Formation 透過增強 AWS Identity and Access Management (IAM) 政策的新授予/撤銷許可模型，提供安全且精細的資料存取。

分析師和資料科學家可以使用 AWS Amazon Athena 等分析和機器學習服務的完整產品組合來存取資料。設定的 Lake Formation 安全政策有助於確保使用者只能存取他們獲授權存取的資料。

混合存取模式

混合存取模式可讓您同時使用 Lake Formation 許可和 IAM 和 Amazon S3 許可來保護和存取目錄資料。混合存取模式可讓資料管理員選擇性地和遞增地加入 Lake Formation 許可，一次專注於一個資料湖使用案例。

藍圖

藍圖是一種資料管理範本，可讓您輕鬆地將資料擷取至資料湖。Lake Formation 提供數個藍圖，每個藍圖適用於預先定義的來源類型，例如關聯式資料庫或 AWS CloudTrail 日誌。從藍圖中，您可以建立工作流程。工作流程由 AWS Glue 爬蟲程式、任務和觸發程序組成，這些觸發程序是用來協調資料的載入和更新。藍圖採用資料來源、資料目標和排程做為輸入，以設定工作流程。

工作流程

工作流程是一組相關AWS Glue任務、爬蟲程式和觸發程序的容器。您可以在 Lake Formation 中建立工作流程，並在AWS Glue服務中執行。Lake Formation 可以將工作流程的狀態追蹤為單一實體。

當您定義工作流程時，您可以選取其基礎的藍圖。然後，您可以隨需或排程執行工作流程。

您在 Lake Formation 中建立的工作流程會在AWS Glue主控台中以定向無環圖 (DAG) 顯示。使用 DAG，您可以追蹤工作流程的進度並執行故障診斷。

Data Catalog

Data Catalog 是您的持久性中繼資料存放區。這是一種受管服務，可讓您以與在 Apache Hive 中繼存放區相同的方式，在 AWS 雲端中存放、標註和共用中繼資料。它提供統一的儲存庫，其中不同的系統可以存放和尋找中繼資料以追蹤資料孤島中的資料，然後使用該中繼資料來查詢和轉換資料。Lake Formation 使用 AWS Glue Data Catalog 來儲存有關資料湖、資料來源、轉換和目標的中繼資料。

有關資料來源和目標的中繼資料採用資料庫和資料表的形式。資料表存放結構描述資訊、位置資訊等。資料庫是資料表的集合。Lake Formation 提供許可階層，以控制對 Data Catalog 中資料庫和資料表的存取。

每個 AWS 帳戶每個 AWS 區域都有一個資料目錄。

基礎資料

基礎資料是指資料目錄資料表指向的資料湖內的來源資料或資料。

Principal

委託人是 AWS Identity and Access Management (IAM) 使用者或角色或 Active Directory 使用者。

資料湖管理員

資料湖管理員是可授予任何委託人（包括自己）任何 Data Catalog 資源或資料位置的任何許可的委託人。將資料湖管理員指定為 Data Catalog 的第一個使用者。然後，此使用者可以將更精細的資源許可授予其他委託人。

> **Note**
>
> IAM 管理使用者 - 具有 `AdministratorAccess` AWS 受管政策的使用者 - 不是自動資料湖管理員。例如，除非已獲授予許可，否則他們無法授予目錄物件的 Lake Formation 許可。不過，他們可以使用 Lake Formation 主控台或 API 將自己指定為資料湖管理員。

如需資料湖管理員功能的資訊，請參閱 隱含 Lake Formation 許可。如需將使用者指定為資料湖管理員的詳細資訊，請參閱 建立資料湖管理員。

AWS 服務與 Lake Formation 整合

您可以使用 Lake Formation 來管理 Amazon S3 中所存放資料的資料庫、資料表和資料欄層級存取許可。向 Lake Formation 註冊資料後，您可以使用 AWS 分析服務 AWS Glue，例如 Amazon Athena、Amazon Redshift Spectrum、Amazon EMR 來查詢資料。下列 AWS 服務與 整合 AWS Lake Formation ，並遵循 Lake Formation 許可。

AWS 服務	整合詳細資訊
AWS Glue	參考主題：AWS Lake Formation 搭配 使用 AWS Glue AWS Glue 和 Lake Formation 共用相同的資料目錄。對於主控台操作（例如檢視資料表清單）和所有 API 操作，AWS Glue使用者只能存取其具有 Lake Formation 許可的資料庫和資料表。
Amazon Athena	參考主題：AWS Lake Formation 搭配 Amazon Athena 使用 使用 Lake Formation 允許或拒絕在 Amazon S3 中讀取資料的許可。當 Amazon Athena 使用者在查詢編輯器中選取AWS Glue目錄時，他們只能查詢其具有 Lake Formation 許可的資料庫、資料表和資料欄。不支援使用資訊清單的查詢。 目前，Lake Formation 不支援管理 Open Table Formats 中 `VACUUM`、`MERGEUPDATE`和 等寫入操作`OPTIMIZE`的許可。 除了透過 AWS Identity and Access Management (IAM) 驗證 Athena 的委託人之外，Lake Formation 還支援透過 JDBC 或 ODBC 驅動程式

AWS 服務	整合詳細資訊
	連線並透過 SAML 驗證的 Athena 使用者。支援的 SAML 供應商包括 Okta 和 Microsoft Active Directory Federation Service (AD FS)。
Amazon Redshift Spectrum	參考主題：AWS Lake Formation 搭配 Amazon Redshift Spectrum 使用 當 Amazon Redshift 使用者在 中的資料庫上建立外部結構描述時 AWS Glue Data Catalog，他們只能查詢該結構描述中具有 Lake Formation 許可的資料表和資料欄。
Amazon QuickSight 企業版	參考： AWS Lake Formation 搭配 QuickSight 使用 當 Amazon QuickSight Enterprise Edition 使用者在 Amazon S3 位置查詢資料集時，使用者必須擁有資料的 Lake Formation SELECT許可。
Amazon EMR	參考： AWS Lake Formation 搭配 Amazon EMR 使用 您可以在建立具有執行期角色的 Amazon EMR 叢集時整合 Lake Formation 許可。 執行時間角色是與 Amazon EMR 任務或查詢建立關聯的 IAM 角色，然後 Amazon EMR 使用此角色來存取 AWS 資源。

Lake Formation 也與 AWS Key Management Service(AWS KMS) 搭配使用，讓您更輕鬆地設定這些整合服務，以加密和解密 Amazon Simple Storage Service (Amazon S3) 位置中的資料。

其他 Lake Formation 資源

如需詳細資訊 AWS Lake Formation，建議您使用下列資源，繼續進一步了解本指南中介紹的概念：

主題

- 部落格
- 技術講座和網路研討會
- 現代架構
- 資料網格資源

- 最佳實務指南

部落格

- AWS Lake Formation 2022 年審核中
- 高彈性的多區域現代化資料架構
- 使用 LF 標籤來引導 IAM 主體的跨帳戶共用
- Lake Formation 許可庫存儀表板
- 事件驅動型資料網格

技術講座和網路研討會

- re：Invent 2020 – 資料湖：輕鬆建置、保護和共用 AWS Lake Formation
- re：Invent 2022 – 在 Amazon S3 上建置和操作資料湖
- AWS Summit SF 2022 – 了解並實現現代資料架構
- AWS Summit ATL 2022 – 使用 AWS Lake Formation、Amazon Redshift 和 的現代資料湖 AWS Glue
- AWS Summit ANZ 2022 – 資料湖、湖房和資料網格：什麼、為什麼和如何？
- AWS 線上技術講座 – 簡化資料湖中的許可和管理

現代架構

- 現代架構模式

資料網格資源

- 使用 AWS Lake Formation 標籤型存取控制大規模建置現代化資料架構和資料網格模式
- JPMorgan Chase 如何建置資料網格架構，以推動顯著價值來增強其企業資料平台
- 在 上建置資料網格 AWS

最佳實務指南

- AWS Lake Formation 最佳實務指南

Lake Formation 入門

我們建議您從下列各節開始著手：

- AWS Lake Formation：運作方式 — 了解基本術語和各種元件的互動方式。
- Lake Formation 入門 — 取得先決條件的相關資訊，並完成重要的設定任務。
- AWS Lake Formation 教學課程 — 遵循step-by-step教學課程，了解如何使用 Lake Formation。
- 中的安全性 AWS Lake Formation — 了解如何協助安全存取 Lake Formation 中的資料。

Lake Formation 入門

如果您尚未註冊 AWS 或需要協助開始使用，請務必完成下列任務。

主題

- 完成初始 AWS 組態任務
- 設定 AWS Lake Formation
- 將AWS Glue資料許可升級到 AWS Lake Formation 模型
- AWS Lake Formation 和介面 VPC 端點 (AWS PrivateLink)

完成初始 AWS 組態任務

若要使用 AWS Lake Formation，您必須先完成以下任務：

主題

- 註冊 AWS 帳戶
- 建立具有管理存取權的使用者
- 授與程式設計存取權

註冊 AWS 帳戶

如果您沒有 AWS 帳戶，請完成下列步驟來建立一個。

註冊 AWS 帳戶

1. 開啟 https://portal.aws.amazon.com/billing/signup。
2. 請遵循線上指示進行。

   部分註冊程序需接收來電，並在電話鍵盤輸入驗證碼。

   當您註冊 時 AWS 帳戶，AWS 帳戶根使用者會建立 。根使用者有權存取該帳戶中的所有 AWS 服務 和資源。作為安全最佳實務，請將管理存取權指派給使用者，並且僅使用根使用者來執行需要根使用者存取權的任務。

AWS 會在註冊程序完成後傳送確認電子郵件給您。您可以隨時登錄 https://aws.amazon.com/ 並選擇我的帳戶，以檢視您目前的帳戶活動並管理帳戶。

建立具有管理存取權的使用者

註冊 後 AWS 帳戶，請保護 AWS 帳戶根使用者、啟用 AWS IAM Identity Center和建立管理使用者，以免將根使用者用於日常任務。

保護您的 AWS 帳戶根使用者

1. 選擇根使用者並輸入 AWS 帳戶 您的電子郵件地址，以帳戶擁有者AWS Management Console身分登入 。在下一頁中，輸入您的密碼。

   如需使用根使用者登入的說明，請參閱 AWS 登入 使用者指南中的以根使用者身分登入。

2. 若要在您的根使用者帳戶上啟用多重要素驗證 (MFA)。

   如需說明，請參閱《IAM 使用者指南》中的為您的 AWS 帳戶 根使用者 （主控台） 啟用虛擬 MFA 裝置。

建立具有管理存取權的使用者

1. 啟用 IAM Identity Center。

   如需指示，請參閱《AWS IAM Identity Center 使用者指南》中的啟用 AWS IAM Identity Center。

2. 在 IAM Identity Center 中，將管理存取權授予使用者。

   如需使用 IAM Identity Center 目錄 做為身分來源的教學課程，請參閱AWS IAM Identity Center《 使用者指南》中的使用預設值設定使用者存取權 IAM Identity Center 目錄。

以具有管理存取權的使用者身分登入

- 若要使用您的 IAM Identity Center 使用者簽署，請使用建立 IAM Identity Center 使用者時傳送至您電子郵件地址的簽署 URL。

  如需使用 IAM Identity Center 使用者登入的說明，請參閱AWS 登入 《 使用者指南》中的登入 AWS 存取入口網站。

指派存取權給其他使用者

1. 在 IAM Identity Center 中，建立一個許可集來遵循套用最低權限的最佳實務。

   如需指示，請參閱《AWS IAM Identity Center 使用者指南》中的建立許可集。
2. 將使用者指派至群組，然後對該群組指派單一登入存取權。

   如需指示，請參閱《AWS IAM Identity Center 使用者指南》中的新增群組。

授與程式設計存取權

如果使用者想要與 AWS 外部互動，則需要程式設計存取 AWS Management Console。授予程式設計存取權的方式取決於正在存取的使用者類型 AWS。

若要授與使用者程式設計存取權，請選擇下列其中一個選項。

哪個使用者需要程式設計存取權？	到	根據
人力資源身分 (IAM Identity Center 中管理的使用者)	使用暫時登入資料簽署對 AWS CLI、 AWS SDKs程式設計請求。 AWS APIs	請依照您要使用的介面所提供的指示操作。 • 如需 AWS CLI，請參閱AWS Command Line Interface 《 使用者指南》中的設定 AWS CLI 要使用 AWS IAM Identity Center的 。 • AWS SDKs、工具和 AWS APIs，請參閱 AWS SDKs 和工具參考指南中的 IAM Identity Center 身分驗證。
IAM	使用暫時登入資料簽署對 AWS CLI、 AWS SDKs程式設計請求。 AWS APIs	遵循《IAM 使用者指南》中將臨時登入資料與 AWS 資源搭配使用中的指示。
IAM	(不建議使用)	請依照您要使用的介面所提供的指示操作。

哪個使用者需要程式設計存取權？	到	根據
	使用長期登入資料來簽署對 AWS CLI、 AWS SDKs程式設計請求。 AWS APIs	• 如需 AWS CLI，請參閱AWS Command Line Interface 《 使用者指南》中的使用 IAM 使用者憑證進行身分驗證。 • AWS SDKs和工具，請參閱 AWS SDKs和工具參考指南中的使用長期憑證進行身分驗證。 • 對於 AWS APIs，請參閱《 IAM 使用者指南》中的管理 IAM 使用者的存取金鑰。

設定 AWS Lake Formation

下列各節提供第一次設定 Lake Formation 的相關資訊。並非本節中的所有主題都需要開始使用 Lake Formation。您可以使用指示來設定 Lake Formation 許可模型，以在 Amazon Simple Storage Service (Amazon S3) 中管理現有的 AWS Glue Data Catalog 物件和資料位置。

1. 建立資料湖管理員
2. 變更預設許可模型或使用混合存取模式
3. the section called “為您的資料湖設定 Amazon S3 位置”
4. the section called “將許可指派給 Lake Formation 使用者”
5. the section called “整合 IAM Identity Center”
6. the section called “（選用）外部資料篩選設定”
7. the section called “（選用）授予對 Data Catalog 加密金鑰的存取權”
8. （選用）建立工作流程的 IAM 角色

本節說明如何以兩種不同的方式設定 Lake Formation 資源：

- 使用 AWS CloudFormation 範本

- 使用 Lake Formation 主控台

若要使用 AWS 主控台設定 Lake Formation，請前往 建立資料湖管理員。

使用 AWS CloudFormation 範本設定 Lake Formation 資源

> **Note**
>
> AWS CloudFormation 堆疊會執行上述的步驟 1 到 6，步驟 2 和 5 除外。從 Lake Formation 主控台執行 變更預設許可模型或使用混合存取模式和 the section called “整合 IAM Identity Center” 手動。

1. 以美國東部（維吉尼亞北部）區域的 IAM 管理員身分登入 AWS CloudFormation 主控台，網址為 https://console.aws.amazon.com/cloudformation：//。
2. 選擇啟動堆疊。
3. 在建立堆疊畫面上選擇下一步。
4. 輸入堆疊名稱。
5. 針對 DatalakeAdminName 和 DatalakeAdminPassword，輸入資料湖管理員使用者的使用者名稱和密碼。
6. 針對 DatalakeUser1Name 和 DatalakeUser1Password，輸入資料湖分析師使用者的使用者名稱和密碼。
7. 針對 DataLakeBucketName，輸入要建立的新儲存貯體名稱。
8. 選擇下一步。
9. 在下一頁中，選擇 `I acknowledge that AWS CloudFormation might create IAM resources with custom names`，然後選擇下一步。
10. 檢閱最終頁面上的詳細資訊，然後選取我確認 AWS CloudFormation 可能會建立 IAM 資源。
11. 選擇建立。

    堆疊建立最多可能需要兩分鐘的時間。

清除資源

如果您想要清除 AWS CloudFormation 堆疊資源：

1. 取消註冊堆疊建立並註冊為資料湖位置的 Amazon S3 儲存貯體。
2. 刪除 AWS CloudFormation 堆疊。這將刪除堆疊建立的所有資源。

建立資料湖管理員

資料湖管理員一開始是唯一 AWS Identity and Access Management (IAM) 使用者或角色，可將資料位置和 Data Catalog 資源的 Lake Formation 許可授予任何主體（包括自己）。如需資料湖管理員功能的詳細資訊，請參閱 隱含 Lake Formation 許可。根據預設，Lake Formation 可讓您建立最多 30 個資料湖管理員。

您可以使用 Lake Formation 主控台或 Lake Formation API `PutDataLakeSettings`的操作來建立資料湖管理員。

建立資料湖管理員需要下列許可。`Administrator` 使用者隱含地擁有這些許可。

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

如果您將`AWSLakeFormationDataAdmin`政策授予使用者，該使用者將無法建立其他 Lake Formation 管理員使用者。

建立資料湖管理員（主控台）

1. 如果要成為資料湖管理員的使用者尚未存在，請使用 IAM 主控台來建立它。否則，請選擇要成為資料湖管理員的現有使用者。

   > **Note**
   >
   > 建議您不要選取 IAM 管理使用者（具有 `AdministratorAccess` AWS 受管政策的使用者）做為資料湖管理員。

   將下列 AWS 受管政策連接至使用者：

政策	強制性？	備註
`AWSLakeFormationDataAdmin`	強制性	基本資料湖管理員許可。此 AWS 受管政策包含 Lake Formation API 操作的明

政策	強制性？	備註
		確拒絕，PutDataLakeSetting 這會限制使用者建立新的資料湖管理員。
AWSGlueConsoleFullAccess，CloudWatchLogsReadOnlyAccess	選用	如果資料湖管理員將疑難排解從 Lake Formation 藍圖建立的工作流程，請連接這些政策。這些政策可讓資料湖管理員在 AWS Glue主控台和 Amazon CloudWatch Logs 主控台中檢視疑難排解資訊。如需工作流程的資訊，請參閱 the section called “使用工作流程匯入資料”。
AWSLakeFormationCrossAccountManager	選用	連接此政策，讓資料湖管理員授予和撤銷 Data Catalog 資源的跨帳戶許可。如需詳細資訊，請參閱Lake Formation 中的跨帳戶資料共用。
AmazonAthenaFullAccess	選用	如果資料湖管理員將在其中執行查詢，請連接此政策 Amazon Athena。

2. 連接下列內嵌政策，授予資料湖管理員建立 Lake Formation 服務連結角色的許可。政策的建議名稱為 LakeFormationSLR。

   服務連結角色可讓資料湖管理員更輕鬆地向 Lake Formation 註冊 Amazon S3 位置。如需 Lake Formation 服務連結角色的詳細資訊，請參閱 the section called “使用服務連結角色”。

   > **Important**
   >
   > 在下列所有政策中，以有效的 AWS 帳號取代 *<account-id>*。

```
{
    "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. （選用）將下列PassRole內嵌政策連接至使用者。此政策可讓資料湖管理員建立和執行工作流程。iam:PassRole 許可可讓工作流程擔任角色LakeFormationWorkflowRole來建立爬蟲程式和任務，並將角色連接到建立的爬蟲程式和任務。政策的建議名稱為 UserPassRole。

> **Important**
>
> 以有效的 AWS 帳號取代 *<account-id>*。

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

4. （選用）如果您的帳戶將授予或接收跨帳戶 Lake Formation 許可，請附加此額外的內嵌政策。此政策可讓資料湖管理員檢視和接受 AWS Resource Access Manager (AWS RAM) 資源共用邀請。此外，對於 AWS Organizations 管理帳戶中的資料湖管理員，該政策包含允許跨帳戶授予組織的許可。如需詳細資訊，請參閱Lake Formation 中的跨帳戶資料共用。

   政策的建議名稱為 RAMAccess。

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

5. 開啟位於 https：//https://console.aws.amazon.com/lakeformation/ 的 AWS Lake Formation 主控台，並以您在 中建立的管理員使用者身分登入，建立具有管理存取權的使用者或以AdministratorAccess使用者 AWS 受管政策的使用者身分登入。
6. 如果出現歡迎使用 Lake Formation 視窗，請選擇您在步驟 1 中建立或選取的 IAM 使用者，然後選擇開始使用。
7. 如果您沒有看到歡迎使用 Lake Formation 視窗，請執行下列步驟來設定 Lake Formation 管理員。
   a. 在導覽窗格中的管理員下，選擇管理角色和任務。在主控台頁面的資料湖管理員區段中，選擇新增。
   b. 在新增管理員對話方塊中的存取類型下，選擇資料湖管理員。
   c. 對於 IAM 使用者和角色，請選擇您在步驟 1 中建立或選取的 IAM 使用者，然後選擇儲存。

變更預設許可模型或使用混合存取模式

Lake Formation 一開始會啟用「僅使用 IAM 存取控制」設定，以與現有 AWS Glue Data Catalog 行為相容。此設定可讓您透過 IAM 政策和 Amazon S3 儲存貯體政策，管理對資料湖中資料及其中繼資料的存取。

為了簡化將資料湖許可從 IAM 和 Amazon S3 模型轉換為 Lake Formation 許可，我們建議您使用 Data Catalog 的混合存取模式。使用混合存取模式時，您會有一個增量路徑，您可以在其中為一組特定使用者啟用 Lake Formation 許可，而不會中斷其他現有的使用者或工作負載。

如需詳細資訊，請參閱混合存取模式。

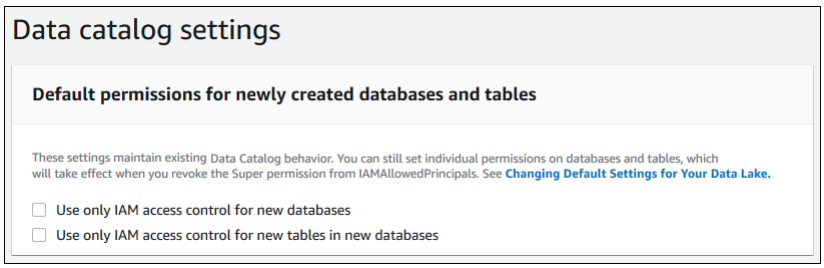
停用預設設定，以在單一步驟中將資料表的所有現有使用者移至 Lake Formation。

> **Important**
>
> 如果您有現有的AWS Glue Data Catalog資料庫和資料表，請勿遵循本節中的指示。或者，請遵循the section called “將AWS Glue資料許可升級至 Lake Formation 模型”中的說明進行。

> **Warning**
>
> 如果您有在 Data Catalog 中建立資料庫和資料表的自動化，下列步驟可能會導致自動化和下游擷取、轉換和載入 (ETL) 任務失敗。只有在您修改現有程序或將明確的 Lake Formation 許可授予必要的委託人之後，才能繼續。如需 Lake Formation 許可的詳細資訊，請參閱 the section called “Lake Formation 許可參考”。

變更預設 Data Catalog 設定

1. 在 Lake Formation 主控台中繼續，網址為 https：//https://console.aws.amazon.com/lakeformation/。確保您以您在 中建立的管理員使用者身分登入，建立具有管理存取權的使用者或以 `AdministratorAccess` AWS 受管政策的使用者身分登入。
2. 修改 Data Catalog 設定：
   a. 在導覽窗格的管理下，選擇 Data Catalog 設定。
   b. 清除兩個核取方塊，然後選擇儲存。



3. 撤銷資料庫建立者的IAMAllowedPrincipals許可。

   a. 在導覽窗格的管理下，選擇管理角色和任務。

   b. 在管理角色和任務主控台頁面的資料庫建立者區段中，選取IAMAllowedPrincipals群組，然後選擇撤銷。

      隨即出現撤銷許可對話方塊，顯示 IAMAllowedPrincipals具有建立資料庫許可。

   c. 選擇撤銷。

將許可指派給 Lake Formation 使用者

建立使用者以存取其中的資料湖 AWS Lake Formation。此使用者具有查詢資料湖的最低權限許可。

如需建立使用者或群組的詳細資訊，請參閱《IAM 使用者指南》中的 IAM 身分。

將存取 Lake Formation 資料的許可連接到非管理員使用者

1. 在 開啟 IAM 主控台，https://console.aws.amazon.com/iam並以您在 中建立的管理員使用者身分登入，建立具有管理存取權的使用者或以 AdministratorAccess AWS 受管政策的使用者身分登入。
2. 選擇使用者或使用者群組。
3. 在清單中，選擇要內嵌政策的使用者或群組名稱。

   選擇許可。
4. 選擇新增許可，然後選擇直接連接政策。Athena 在篩選政策文字欄位中輸入 。在結果清單中，勾選 的方塊AmazonAthenaFullAccess。
5. 選擇建立政策按鈕。在建立政策頁面上，選擇 JSON 標籤。將下列程式碼複製並貼到政策編輯器中。

```
{
    "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. 選擇底部的下一步按鈕，直到您看到檢閱政策頁面為止。輸入政策的名稱，例如 `DatalakeUserBasic`。選擇建立政策，然後關閉政策索引標籤或瀏覽器視窗。

為您的資料湖設定 Amazon S3 位置

若要使用 Lake Formation 來管理和保護資料湖中的資料，您必須先註冊 Amazon S3 位置。當您註冊位置時，Amazon S3 路徑和該路徑下的所有資料夾都會註冊，這可讓 Lake Formation 強制執行儲存層級許可。當使用者從 Amazon Athena 等整合引擎請求資料時，Lake Formation 會提供資料存取，而不是使用使用者許可。

註冊位置時，您可以指定 IAM 角色，授予該位置的讀取/寫入許可。Lake Formation 在將臨時登入資料提供給請求存取已註冊 Amazon S3 位置中資料的整合 AWS 服務時，會擔任該角色。您可以指定 Lake Formation 服務連結角色 (SLR) 或建立自己的角色。

在下列情況下使用自訂角色：

- 您計劃在 Amazon CloudWatch Logs 中發佈指標。使用者定義的角色必須包含在 CloudWatch Logs 中新增日誌，以及除了 SLR 許可之外發佈指標的政策。如需授予必要 CloudWatch 許可的內嵌政策範例，請參閱 用於註冊位置的角色需求。

- Amazon S3 位置存在於不同的帳戶中。如需詳細資訊，請參閱the section called “在另一個 AWS 帳戶中註冊 Amazon S3 位置”。
- Amazon S3 位置包含使用 加密的資料 AWS 受管金鑰。如需詳細資訊，請參閱 註冊加密的 Amazon S3 位置 和 跨 AWS 帳戶註冊加密的 Amazon S3 位置。
- 您計劃使用 Amazon EMR 存取 Amazon S3 位置。如需角色需求的詳細資訊，請參閱《Amazon EMR 管理指南》中的 Lake Formation 的 IAM 角色。

您選擇的角色必須具有必要的許可，如中所述用於註冊位置的角色需求。如需如何註冊 Amazon S3 位置的說明，請參閱 將 Amazon S3 位置新增至您的資料湖。

（選用）外部資料篩選設定

如果您想要使用第三方查詢引擎來分析和處理資料湖中的資料，您必須選擇加入，以允許外部引擎存取 Lake Formation 管理的資料。如果您不選擇加入，外部引擎將無法存取在 Lake Formation 註冊的 Amazon S3 位置中的資料。

Lake Formation 支援資料欄層級許可，以限制對資料表中特定資料欄的存取。整合式分析服務 Amazon Athena，例如 Amazon Redshift Spectrum 和 Amazon EMR 從 擷取未篩選的資料表中繼資料 AWS Glue Data Catalog。查詢回應中資料欄的實際篩選是整合服務的責任。第三方管理員有責任正確處理許可，以避免未經授權的資料存取。

選擇加入以允許第三方引擎存取和篩選資料（主控台）

1. 在 Lake Formation 主控台中繼續，網址為 https：//https://console.aws.amazon.com/lakeformation/。確保您以擁有 Lake Formation `PutDataLakeSettings` API 操作 IAM 許可的委託人身分登入。您在 中建立的 IAM 管理員使用者註冊 AWS 帳戶具有此許可。
2. 在導覽窗格的管理下，選擇應用程式整合設定。
3. 在應用程式整合設定頁面上，執行下列動作：
   a. 勾選方塊 允許外部引擎篩選向 Lake Formation 註冊的 Amazon S3 位置中的資料。
   b. 輸入為第三方引擎定義的工作階段標籤值。
   c. 針對AWS 帳戶 IDs，輸入允許第三方引擎存取向 Lake Formation 註冊之位置的帳戶 IDs。在每個帳戶 ID 之後按 Enter。
   d. 選擇儲存。

若要允許外部引擎在沒有工作階段標籤驗證的情況下存取資料，請參閱 完整資料表存取的應用程式整合

（選用）授予對 Data Catalog 加密金鑰的存取權

如果 AWS Glue Data Catalog 已加密，請將 AWS KMS 金鑰上的 AWS Identity and Access Management 許可授予需要對 Data Catalog 資料庫和資料表授予 Lake Formation 許可的任何委託人。

如需詳細資訊，請參閱《AWS Key Management Service 開發人員指南》。

（選用）建立工作流程的 IAM 角色

透過 AWS Lake Formation，您可以使用 AWS Glue 爬蟲程式執行的工作流程匯入資料。工作流程會定義資料來源和排程，將資料匯入您的資料湖。您可以使用 Lake Formation 提供的藍圖或範本輕鬆定義工作流程。

建立工作流程時，您必須為其指派 AWS Identity and Access Management (IAM) 角色，授予 Lake Formation 擷取資料的必要許可。

下列程序假設熟悉 IAM。

為工作流程建立 IAM 角色

1. 在 開啟 IAM 主控台，https://console.aws.amazon.com/iam並以您在 中建立的管理員使用者身分登入，建立具有管理存取權的使用者或以 `AdministratorAccess` AWS 受管政策的使用者身分登入。
2. 在導覽窗格中，選擇角色，然後選擇建立角色。
3. 在建立角色頁面上，選擇AWS 服務，然後選擇 Glue。選擇下一步。
4. 在新增許可頁面上，搜尋 AWSGlueServiceRole 受管政策，然後選取清單中政策名稱旁的核取方塊。然後完成建立角色精靈，命名角色 `LFWorkflowRole`。若要完成，請選擇建立角色。
5. 返回角色頁面，搜尋 `LFWorkflowRole`，然後選擇角色名稱。
6. 在角色摘要頁面的許可索引標籤下，選擇建立內嵌政策。在建立政策畫面上，導覽至 JSON 索引標籤，然後新增下列內嵌政策。政策的建議名稱為 `LakeFormationWorkflow`。

> **Important**
>
> 在下列政策中，將 *<account-id>* 取代為有效的 AWS 帳戶 數字。

```
{
    "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"
            ]
        }
    ]
}
```

以下是此政策中許可的簡短描述：

- `lakeformation:GetDataAccess` 可讓工作流程建立的任務寫入目標位置。
- `lakeformation:GrantPermissions` 可讓工作流程授予目標資料表的SELECT許可。
- `iam:PassRole` 可讓服務擔任角色LakeFormationWorkflowRole以建立爬蟲程式和任務（工作流程執行個體），並將角色連接至建立的爬蟲程式和任務。

7. 確認角色LakeFormationWorkflowRole已連接兩個政策。
8. 如果您要擷取資料湖位置以外的資料，請新增內嵌政策，授予讀取來源資料的許可。

將AWS Glue資料許可升級到 AWS Lake Formation 模型

AWS Lake Formation 許可可針對資料湖中的資料啟用精細存取控制。您可以使用 Lake Formation 許可模型，在 Amazon Simple Storage Service (Amazon S3) 中管理現有的 AWS Glue Data Catalog 物件和資料位置。

Lake Formation 許可模型使用粗粒 AWS Identity and Access Management (IAM) 許可進行 API 服務存取。Lake Formation 使用 Lake Formation 中的資料篩選和儲存格層級安全性功能來限制使用者及其應用程式在資料欄、資料列和儲存格層級的資料表存取。相較之下，AWS Glue模型會透過以身分為基礎的和資源為基礎的 IAM 政策授予資料存取權。

若要進行切換，請遵循本指南中的步驟。

如需詳細資訊，請參閱Lake Formation 許可概觀 。

關於預設許可

若要維持與 的回溯相容性AWS Glue，預設會將 Super許可 AWS Lake Formation 授予所有現有 AWS Glue Data Catalog 資源上的 IAMAllowedPrincipals群組，並在啟用僅使用 IAM 存取控制設定時授予新 Data Catalog 資源的 Super許可。這實際上會導致僅由 AWS Identity and Access Management (IAM) 政策控制對 Data Catalog 資源和 Amazon S3 位置的存取。IAMAllowedPrincipals 群組包含 IAM 政策允許存取 Data Catalog 物件的任何 IAM 使用者和角色。Super 許可可讓委託人在授予其資料庫或資料表上執行每個支援的 Lake Formation 操作。

您可以透過在 Lake Formation 中註冊現有 Data Catalog 資源的位置或使用混合存取模式，開始使用 Lake Formation 來管理對資料的存取。在混合存取模式中註冊 Amazon S3 位置時，您可以選擇該位置下資料庫和資料表的主體，以啟用 Lake Formation 許可。

為了簡化將資料湖許可從 IAM 和 Amazon S3 模型轉換為 Lake Formation 許可，我們建議您使用 Data Catalog 的混合存取模式。使用混合存取模式時，您會有一個增量路徑，您可以在其中為特定一組使用者啟用 Lake Formation 許可，而不會中斷其他現有使用者或工作負載。

如需詳細資訊，請參閱混合存取模式。

停用預設 Data Catalog 設定，在單一步驟中將資料表的所有現有使用者移至 Lake Formation。

若要開始將 Lake Formation 許可與現有的 AWS Glue Data Catalog 資料庫和資料表搭配使用，您必須執行下列動作：

1. 判斷使用者對每個資料庫和資料表的現有 IAM 許可。
2. 在 Lake Formation 中複寫這些許可。
3. 對於包含資料的每個 Amazon S3 位置：
   a. 從參考該位置的每個 Data Catalog 資源上的 IAMAllowedPrincipals群組撤銷Super許可。
   b. 向 Lake Formation 註冊位置。

4. 清除現有的精細存取控制 IAM 政策。

> **Important**
>
> 若要在轉換 Data Catalog 的過程中新增使用者，您必須在 IAM 中像之前一樣設定精細的 AWS Glue許可。您也必須在 Lake Formation 中複寫這些許可，如本節所述。如果新使用者具有本指南中所述的粗略 IAM 政策，他們可以列出具有授予 Super許可的任何資料庫或資料表IAMAllowedPrincipals。他們也可以檢視這些資源的中繼資料。

遵循本節中的步驟，升級至 Lake Formation 許可模型。

主題

- 步驟 1：列出使用者和角色的現有許可
- 步驟 2：設定對等的 Lake Formation 許可
- 步驟 3：授予使用者使用 Lake Formation 的 IAM 許可
- 步驟 4：將資料存放區切換至 Lake Formation 許可模型
- 步驟 5：保護新的 Data Catalog 資源
- 步驟 6：為使用者提供新的 IAM 政策，以供未來存取資料湖
- 步驟 7：清除現有的 IAM 政策

步驟 1：列出使用者和角色的現有許可

若要開始對現有的AWS Glue資料庫和資料表使用 AWS Lake Formation 許可，您必須先判斷使用者的現有許可。

> **Important**
>
> 開始之前，請確定您已完成 中的任務開始使用。

主題

- 使用 API 操作
- 使用 AWS Management Console
- 使用 AWS CloudTrail

使用 API 操作

使用 AWS Identity and Access Management (IAM) ListPoliciesGrantingServiceAccess API 操作來判斷連接到每個委託人（使用者或角色）的 IAM 政策。從結果中傳回的政策，您可以判斷授予委託人的 IAM 許可。您必須分別叫用每個委託人的 API。

Example

下列 AWS CLI 範例會傳回連接到使用者 的政策`glue_user1`。

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

命令會傳回類似下列的結果。

```
{
    "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
}
```

使用 AWS Management Console

您也可以在 AWS Identity and Access Management (IAM) 主控台、使用者或角色摘要頁面上的 Access Advisor 索引標籤中查看此資訊：

1. 前往 https://console.aws.amazon.com/iam/ 開啟 IAM 主控台。
2. 在服務導覽窗格中，選擇 Users (使用者) 或者 Roles (角色)。
3. 在清單中選擇名稱以開啟其摘要頁面，然後選擇 Access Advisor 索引標籤。
4. 檢查每個政策，以判斷每個使用者具有許可的資料庫、資料表和動作組合。

   請記得在此過程中檢查使用者以外的角色，因為資料處理任務可能擔任角色來存取資料。

使用 AWS CloudTrail

判斷您現有許可的另一種方法是尋找日誌 `additionaleventdata` 欄位包含 `insufficientLakeFormationPermissions`項目的 AWS CloudTrail AWS Glue API 呼叫。此項目會列出使用者需要 Lake Formation 許可才能採取相同動作的資料庫和資料表。

這些是資料存取日誌，因此無法保證它們會產生完整的使用者及其許可清單。建議您選擇廣泛的時間範圍來擷取大部分使用者的資料存取模式，例如數週或數月。

如需詳細資訊，請參閱《AWS CloudTrail 使用者指南》中的使用 CloudTrail 事件歷史記錄檢視事件。

接下來，您可以設定 Lake Formation 許可以符合AWS Glue許可。請參閱 步驟 2：設定對等的 Lake Formation 許可。

步驟 2：設定對等的 Lake Formation 許可

使用 中收集的資訊步驟 1：列出使用者和角色的現有許可，授予許可 AWS Lake Formation 以符合 AWS Glue許可。使用下列任一方法來執行授予：

- 使用 Lake Formation 主控台或 AWS CLI。

  請參閱 the section called “授予資料許可”。
- 使用 `GrantPermissions`或 `BatchGrantPermissions` API 操作。

  請參閱 許可 APIs。

如需詳細資訊，請參閱Lake Formation 許可概觀 。

設定 Lake Formation 許可後，請繼續 步驟 3：授予使用者使用 Lake Formation 的 IAM 許可。

步驟 3：授予使用者使用 Lake Formation 的 IAM 許可

若要使用 AWS Lake Formation 許可模型，委託人必須擁有 Lake Formation APIs的 AWS Identity and Access Management (IAM) 許可。

在 IAM 中建立下列政策，並將其連接到需要存取資料湖的每個使用者。將政策命名為 `LakeFormationDataAccess`。

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

接下來，一次升級到一個資料位置的 Lake Formation 許可。請參閱 步驟 4：將資料存放區切換至 Lake Formation 許可模型。

步驟 4：將資料存放區切換至 Lake Formation 許可模型

一次升級至 Lake Formation 許可一個資料位置。若要這樣做，請重複整個區段，直到您已註冊資料目錄參考的所有 Amazon Simple Storage Service (Amazon S3) 路徑。

主題

- 驗證 Lake Formation 許可
- 保護現有的 Data Catalog 資源
- 開啟 Amazon S3 位置的 Lake Formation 許可

驗證 Lake Formation 許可

註冊位置之前，請執行驗證步驟，以確保正確的主體具有所需的 Lake Formation 許可，而且不會將 Lake Formation 許可授予不應擁有它們的主體。使用 Lake Formation

GetEffectivePermissionsForPath API 操作，識別參考 Amazon S3 位置的資料目錄資源，以及具有這些資源許可的主體。

下列 AWS CLI 範例會傳回參考 Amazon S3 儲存貯體 的 Data Catalog 資料庫和資料表products。

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

請注意 profile選項。建議您以資料湖管理員身分執行 命令。

以下是傳回結果的摘錄。

```
{
        "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**
>
> 如果您的 AWS Glue Data Catalog 已加密，則 只會GetEffectivePermissionsForPath傳回在 Lake Formation 一般可用性之後建立或修改的資料庫和資料表。

保護現有的 Data Catalog 資源

接著，撤銷您為位置識別的每個資料表和資料庫`IAMAllowedPrincipals`的 `Super`許可。

> **Warning**
>
> 如果您有在 Data Catalog 中建立資料庫和資料表的自動化，下列步驟可能會導致自動化和下游擷取、轉換和載入 (ETL) 任務失敗。只有在您修改現有程序或將明確的 Lake Formation 許可授予必要的委託人之後，才能繼續執行。如需 Lake Formation 許可的詳細資訊，請參閱 the section called “Lake Formation 許可參考”。

在資料表**`IAMAllowedPrincipals`**上**`Super`**撤銷

1. 開啟 AWS Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/：//。以資料湖管理員身分登入。
2. 在導覽窗格中，選擇 Tables (資料表)。
3. 在資料表頁面上，選取所需資料表旁的選項按鈕。
4. 在動作功能表中，選擇撤銷。
5. 在撤銷許可對話方塊中，在 IAM 使用者和角色清單中，向下捲動至群組標題，然後選擇 IAMAllowedPrincipals。
6. 在資料表許可下，確定已選取超級，然後選擇撤銷。

在資料庫**`IAMAllowedPrincipals`**上**`Super`**撤銷

1. 開啟 AWS Lake Formation 主控台，網址為 https：//https://console.aws.amazon.com/lakeformation/。以資料湖管理員身分登入。
2. 在導覽窗格中，選擇 Databases (資料庫)。
3. 在資料庫頁面上，選取所需資料庫旁的選項按鈕。
4. 在 Actions (動作) 功能表上，選擇 Edit (編輯)。
5. 在編輯資料庫頁面上，清除僅對此資料庫中的新資料表使用 IAM 存取控制，然後選擇儲存。
6. 返回資料庫頁面，確保資料庫仍處於選取狀態，然後在動作功能表上，選擇撤銷。
7. 在撤銷許可對話方塊中，在 IAM 使用者和角色清單中，向下捲動至群組標題，然後選擇 IAMAllowedPrincipals。
8. 在資料庫許可下，確定已選取超級，然後選擇撤銷。

開啟 Amazon S3 位置的 Lake Formation 許可

接著，向 Lake Formation 註冊 Amazon S3 位置。若要這樣做，您可以使用中所述的程序將 Amazon S3 位置新增至您的資料湖。或者，使用 `RegisterResource` API 操作，如 中所述登入資料販賣 APIs。

> **Note**
>
> 如果父位置已註冊，則不需要註冊子位置。

完成這些步驟並測試使用者是否可以存取其資料後，您已成功升級至 Lake Formation 許可。繼續下一個步驟：步驟 5：保護新的 Data Catalog 資源。

步驟 5：保護新的 Data Catalog 資源

接下來，變更預設 Data Catalog 設定來保護所有新的 Data Catalog 資源。關閉選項以僅對新資料庫和資料表使用 AWS Identity and Access Management (IAM) 存取控制。

> **Warning**
>
> 如果您有在 Data Catalog 中建立資料庫和資料表的自動化，下列步驟可能會導致自動化和下游擷取、轉換和載入 (ETL) 任務失敗。只有在您修改現有程序或將明確的 Lake Formation 許可授予必要的委託人之後，才能繼續執行。如需 Lake Formation 許可的詳細資訊，請參閱 the section called “Lake Formation 許可參考”。

變更預設 Data Catalog 設定

1. 開啟 AWS Lake Formation 主控台，網址為 https：//https://console.aws.amazon.com/lakeformation/。以 IAM 管理使用者 （ 使用者`Administrator`或其他具有 `AdministratorAccess` AWS 受管政策的使用者 ） 身分登入。
2. 在導覽窗格中，選擇設定。
3. 在資料目錄設定頁面上，清除兩個核取方塊，然後選擇儲存。

下一步是授予使用者未來存取其他資料庫或資料表的權限。請參閱 步驟 6：為使用者提供新的 IAM 政策，以供未來存取資料湖。

步驟 6：為使用者提供新的 IAM 政策，以供未來存取資料湖

若要授予使用者未來對其他 Data Catalog 資料庫或資料表的存取權，您必須為他們提供以下粗略粒狀 AWS Identity and Access Management (IAM) 內嵌政策。將政策命名為 `GlueFullReadAccess`。

> **Important**
>
> 如果您在`Super`資料目錄中的每個資料庫和資料表`IAMAllowedPrincipals`上從 撤銷之前，將此政策連接到使用者，則該使用者可以檢視 `Super` 上授予 的任何資源的所有中繼資料`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**
>
> 此步驟和先前步驟中指定的內嵌政策包含最少的 IAM 許可。如需資料湖管理員、資料分析師和其他角色的建議政策，請參閱 the section called “Lake Formation 角色和 IAM 許可參考”。

接著，繼續 步驟 7：清除現有的 IAM 政策。

步驟 7：清除現有的 IAM 政策

設定 AWS Lake Formation 許可並建立並連接粗略存取控制 AWS Identity and Access Management (IAM) 政策後，請完成下列最後步驟：

- 從使用者、群組和角色中移除您在 Lake Formation 中複寫的舊精細存取控制 IAM 政策。

如此一來，您就能確保這些委託人無法再直接存取 Amazon Simple Storage Service (Amazon S3) 中的資料。然後，您可以完全透過 Lake Formation 管理這些委託人的資料湖存取。

AWS Lake Formation 和介面 VPC 端點 (AWS PrivateLink)

Amazon VPC 是一項 AWS 服務，可用來在您定義的虛擬網路中啟動 AWS 資源。您可利用 VPC 來控制您的網路設定，例如 IP 地址範圍、子網路、路由表和網路閘道。

如果您使用 Amazon Virtual Private Cloud (Amazon VPC) 託管 AWS 資源，您可以在 VPC 和 Lake Formation 之間建立私有連線。您可以使用此連線，讓 Lake Formation 可以與 VPC 中的資源通訊，而無需透過公有網際網路。

您可以建立介面 VPC 端點， AWS Lake Formation 在 VPC 和 之間建立私有連線。 介面端點採用 AWS PrivateLink技術，可讓您在沒有網際網路閘道、NAT 裝置、VPN 連線或 AWS Direct Connect 連線的情況下，私下存取 Lake Formation APIs。VPC 中的執行個體不需要公有 IP 地址，即可與 Lake Formation APIs通訊。VPC 和 Lake Formation 之間的流量不會離開 Amazon 網路。

每個介面端點都是由您子網路中的一或多個彈性網路介面表示。

如需詳細資訊，請參閱《Amazon VPC 使用者指南》中的介面 VPC 端點 (AWS PrivateLink)。

Lake Formation VPC 端點的考量事項

設定 Lake Formation 的介面 VPC 端點之前，請務必檢閱《Amazon VPC 使用者指南》中的介面端點屬性和限制。

Lake Formation 支援從您的 VPC 呼叫其所有 API 動作。您可以在 AWS 區域 支援 Lake Formation 和 Amazon VPC 端點的所有 中使用 Lake Formation 搭配 VPC 端點。

為 Lake Formation 建立介面 VPC 端點

您可以使用 Amazon VPC 主控台或 AWS Command Line Interface () 為 Lake Formation 服務建立 VPC 端點AWS CLI。如需詳細資訊，請參閱《Amazon VPC 使用者指南》中的建立介面端點。

使用下列服務名稱建立 Lake Formation 的 VPC 端點：

- com.amazonaws.*region*.lakeformation

如果您為端點啟用私有 DNS，則可以使用區域的預設 DNS 名稱向 Lake Formation 提出 API 請求，例如 `lakeformation.us-east-1.amazonaws.com`。

如需詳細資訊，請參閱《Amazon VPC 使用者指南》中的透過介面端點存取服務。

為 Lake Formation 建立 VPC 端點政策

Lake Formation 支援 VPC 端點政策。端點政策是一種以資源為基礎的政策，您可以連接到 VPC 端點，以控制哪些 AWS 主體可以使用端點來存取 AWS 服務。

您可以將端點政策連接至控制 Lake Formation 存取的 VPC 端點。此政策會指定下列資訊：

- 可執行動作的主體。
- 可執行的動作。
- 可供執行動作的資源。

如需詳細資訊，請參閱 Amazon VPC 使用者指南中的使用 VPC 端點控制對服務的存取。

範例：Lake Formation 動作的 VPC 端點政策

下列 Lake Formation 的 VPC 端點政策範例允許使用 Lake Formation 許可來販賣登入資料。您可以使用此政策，從 Amazon Redshift 叢集或私有子網路中的 Amazon EMR 叢集使用 Lake Formation 許可來執行查詢。

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

> **Note**
>
> 如果您在建立端點時未連接政策，則會連接允許完整存取服務的預設政策。

如需詳細資訊，請參閱 Amazon VPC 文件中的下列主題：

- 什麼是 Amazon VPC？
- 建立界面端點
- 使用 VPC 端點政策

AWS Lake Formation 教學課程

下列教學課程分為三個軌道，並提供逐步 step-by-step 指示，說明如何使用 建置資料湖、擷取資料、共用和保護資料湖 AWS Lake Formation：

1. 建置資料湖並擷取資料：了解如何建置資料湖，並使用藍圖來移動、存放、編目、清理和整理資料。您也將學習如何設定受管資料表。受管資料表是一種新的 Amazon S3 資料表類型，支援原子、一致、隔離和耐久 (ACID) 交易。

   開始之前，請確定您已完成 中的步驟Lake Formation 入門。

   - 從 AWS CloudTrail 來源建立資料湖

     使用您自己的 CloudTrail 日誌做為資料來源，建立和載入您的第一個資料湖。

   - 從 Lake Formation 中的 JDBC 來源建立資料湖

     使用其中一個可存取 JDBC 的資料存放區建立資料湖，例如關聯式資料庫，做為資料來源。

2. 保護資料湖：了解如何使用標籤型和資料列層級存取控制，以有效保護和管理對資料湖的存取。

   - 在 Lake Formation 中設定開放資料表儲存格式的許可

     本教學課程示範如何在 Lake Formation 中設定開放原始碼交易資料表格式 (Apache Iceberg、Apache Hudi 和 Linux Foundation Delta Lake 資料表） 的許可。

   - 使用 Lake Formation 標籤型存取控制管理資料湖

     了解如何使用 Lake Formation 中的標籤型存取控制，管理資料湖內資料的存取。

   - 使用資料列層級存取控制保護資料湖

     了解如何設定資料列層級許可，讓您根據 Lake Formation 中的資料合規和控管政策，限制對特定資料列的存取。

3. 共用資料：了解如何 AWS 帳戶 使用標籤型存取控制 (TBAC) 安全地跨 共用資料，並管理共用資料集的精細許可 AWS 帳戶。

   - 使用 Lake Formation 標籤型存取控制和具名資源共用資料湖

     在本教學課程中，您將了解如何使用 Lake Formation 安全地跨 AWS 帳戶 共用資料。

   - 使用 Lake Formation 精細存取控制共用資料湖

     在本教學課程中，您將了解如何在 AWS 帳戶 管理多個 時，使用 Lake Formation 快速輕鬆地共用資料集 AWS Organizations。

主題

- 從 AWS CloudTrail 來源建立資料湖
- 從 Lake Formation 中的 JDBC 來源建立資料湖
- 在 Lake Formation 中設定開放資料表儲存格式的許可
- 使用 Lake Formation 標籤型存取控制管理資料湖
- 使用資料列層級存取控制保護資料湖
- 使用 Lake Formation 標籤型存取控制和具名資源共用資料湖
- 使用 Lake Formation 精細存取控制共用資料湖

從 AWS CloudTrail 來源建立資料湖

本教學課程會引導您完成在 Lake Formation 主控台上要採取的動作，以從 AWS CloudTrail 來源建立和載入您的第一個資料湖。

建立資料湖的高階步驟

1. 將 Amazon Simple Storage Service (Amazon S3) 路徑註冊為資料湖。
2. 授予 Lake Formation 許可，以寫入資料目錄和資料湖中的 Amazon S3 位置。
3. 建立資料庫以組織 Data Catalog 中的中繼資料表。
4. 使用藍圖來建立工作流程。執行工作流程以從資料來源擷取資料。
5. 設定 Lake Formation 許可，以允許其他人管理 Data Catalog 和資料湖中的資料。
6. 設定 Amazon Athena 來查詢您匯入 Amazon S3 資料湖的資料。
7. 對於某些資料存放區類型，請設定 Amazon Redshift Spectrum 來查詢您匯入 Amazon S3 資料湖的資料。

主題

- 目標對象
- 先決條件
- 步驟 1：建立資料分析師使用者
- 步驟 2：將讀取 AWS CloudTrail 日誌的許可新增至工作流程角色
- 步驟 3：為資料湖建立 Amazon S3 儲存貯體
- 步驟 4：註冊 Amazon S3 路徑

- 步驟 5：授予資料位置許可
- 步驟 6：在 Data Catalog 中建立資料庫
- 步驟 7：授予資料許可
- 步驟 8：使用藍圖建立工作流程
- 步驟 9：執行工作流程
- 步驟 10：在資料表上授予 SELECT
- 步驟 11：使用 查詢資料湖 Amazon Athena

目標對象

下表列出本教學課程中用於建立資料湖的角色。

目標對象

角色	描述
IAM 管理員	具有 AWS 受管政策：`AdministratorAccess` 。可以建立 IAM 角色和 Amazon S3 儲存貯體。
資料湖管理員	可以存取資料目錄、建立資料庫，以及將 Lake Formation 許可授予其他使用者的使用者。IAM 許可少於 IAM 管理員，但足以管理資料湖。
資料分析	可對資料湖執行查詢的使用者。僅有足夠許可來執行查詢。
工作流程角色	具有執行工作流程所需 IAM 政策的角色。如需詳細資訊，請參閱（選用）建立工作流程的 IAM 角色。

先決條件

開始之前：

- 請確定您已完成 中的任務設定 AWS Lake Formation。
- 了解 CloudTrail 日誌的位置。

- Athena 需要資料分析師角色來建立 Amazon S3 儲存貯體，以便在使用 Athena 之前存放查詢結果。

假設熟悉 AWS Identity and Access Management (IAM)。如需 IAM 的相關資訊，請參閱 IAM 使用者指南。

步驟 1：建立資料分析師使用者

此使用者具有查詢資料湖的最小許可集。

1. 在 https://console.aws.amazon.com/iam 開啟 IAM 主控台。以您在 中建立的管理員使用者身分登入，建立具有管理存取權的使用者或以 AdministratorAccess AWS 受管政策的使用者身分登入。
2. datalake_user 使用下列設定建立名為 的使用者：
   - 啟用 AWS Management Console 存取。
   - 設定密碼，不需要重設密碼。
   - 連接 AmazonAthenaFullAccess AWS 受管政策。
   - 連接下列內嵌政策。將政策命名為 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": "*"
        }
```

```
    ]
}
```

步驟 2：將讀取 AWS CloudTrail 日誌的許可新增至工作流程角色

1. 將下列內嵌政策連接至角色 LakeFormationWorkflowRole。政策會授予讀取 AWS CloudTrail 日誌的許可。將政策命名為 DatalakeGetCloudTrail。

   若要建立 LakeFormationWorkflowRole 角色，請參閱（選用）建立工作流程的 IAM 角色。

   > **Important**
   >
   > 將 *<your-s3-cloudtrail-bucket>* 取代為 CloudTrail 資料的 Amazon S3 位置。

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

2. 確認有三個政策連接到角色。

步驟 3：為資料湖建立 Amazon S3 儲存貯體

建立要作為資料湖根位置的 Amazon S3 儲存貯體。

1. 開啟 Amazon S3 主控台，網址為 https://console.aws.amazon.com/s3/：//建立具有管理存取權的使用者https：//www.healthnet.com/healthnet.com/healthnet.com/healthnet.com/healthnet.com/healthnet.com/healthnet.com/healthnet.com/com；Amazon S3 主控台，
2. 選擇建立儲存貯體，然後瀏覽精靈以建立名為 的儲存貯體*<yourName>*-datalake-cloudtrail，其中 *<yourName>* 是您的第一個初始名稱和姓氏。例如：jdoe-datalake-cloudtrail。

如需建立 Amazon S3 儲存貯體的詳細說明，請參閱建立儲存貯體。

步驟 4：註冊 Amazon S3 路徑

將 Amazon S3 路徑註冊為資料湖的根位置。

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。以資料湖管理員身分登入。
2. 在導覽窗格的註冊和擷取下，選擇資料湖位置。
3. 選擇註冊位置，然後瀏覽。
4. 選取您先前建立的儲存*<yourName>*-datalake-cloudtrail貯體，接受預設的 IAM 角色 AWSServiceRoleForLakeFormationDataAccess，然後選擇註冊位置。

   如需註冊位置的詳細資訊，請參閱 將 Amazon S3 位置新增至您的資料湖。

步驟 5：授予資料位置許可

主體必須在資料湖位置上擁有資料位置許可，才能建立指向該位置的資料目錄資料表或資料庫。您必須將資料位置許可授予工作流程的 IAM 角色，工作流程才能寫入資料擷取目的地。

1. 在導覽窗格中的許可下，選擇資料位置。
2. 選擇授予，然後在授予許可對話方塊中進行這些選擇：
   a. 針對 IAM 使用者和角色，選擇 LakeFormationWorkflowRole。
   b. 針對儲存位置，選擇您的儲存*<yourName>*-datalake-cloudtrail貯體。
3. 選擇 Grant (授予)。

如需資料位置許可的詳細資訊，請參閱 Underlying data access control。

步驟 6：在 Data Catalog 中建立資料庫

Lake Formation Data Catalog 中的中繼資料表會存放在資料庫中。

1. 在導覽窗格中的資料目錄下，選擇資料庫。
2. 選擇建立資料庫，然後在資料庫詳細資訊下，輸入名稱 lakeformation_cloudtrail。
3. 將其他欄位保留空白，然後選擇建立資料庫。

步驟 7：授予資料許可

您必須授予在 Data Catalog 中建立中繼資料表的許可。由於工作流程將執行角色 `LakeFormationWorkflowRole`，您必須將這些許可授予角色。

1. 在 Lake Formation 主控台的導覽窗格中的資料目錄下，選擇資料庫。
2. 選擇`lakeformation_cloudtrail`資料庫，然後從動作下拉式清單中，選擇許可標題下的授予。
3. 在授予資料許可對話方塊中，進行下列選擇：
   a. 在主體下，針對 IAM 使用者和角色，選擇 `LakeFormationWorkflowRole`。
   b. 在 LF-標籤或型錄資源下，選擇已命名的的資料型錄資源。
   c. 對於資料庫，您應該會看到`lakeformation_cloudtrail`資料庫已新增。
   d. 在資料庫許可下，選取建立資料表、修改和捨棄，並在選取時清除超級。
4. 選擇 Grant (授予)。

如需授予 Lake Formation 許可的詳細資訊，請參閱 管理 Lake Formation 許可。

步驟 8：使用藍圖建立工作流程

為了讀取 CloudTrail 日誌、了解其結構、在 Data Catalog 中建立適當的資料表，我們需要設定由AWS Glue爬蟲程式、任務、觸發條件和工作流程組成的工作流程。Lake Formation 的藍圖簡化了此程序。

工作流程會產生任務、爬蟲程式和觸發條件，以探索資料並將其擷取至您的資料湖。您可以根據其中一個預先定義的 Lake Formation 藍圖建立工作流程。

1. 在 Lake Formation 主控台的導覽窗格中，選擇擷取下的藍圖，然後選擇使用藍圖。
2. 在使用藍圖頁面的藍圖類型下，選擇 AWS CloudTrail。
3. 在匯入來源下，選擇 CloudTrail 來源和開始日期。
4. 在匯入目標下，指定下列參數：

目標資料庫	`lakeformation_cloudtrail`
目標儲存位置	`s3://<yourName> -datalake-cloudtrail`

資料格式	Parquet

5. 針對匯入頻率，選擇隨需執行。
6. 在匯入選項下，指定下列參數：

工作流程名稱	`lakeformationcloudtrailtest`
IAM 角色	`LakeFormationWorkflowRole`
資料表字首	`cloudtrailtest` **Note** 必須為小寫。

7. 選擇建立，然後等待主控台報告工作流程已成功建立。

> **Tip**
> 您是否收到下列錯誤訊息？
> `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...`
> 若是如此，請檢查您是否已將資料湖管理員使用者的內嵌政策中的 *<account-id>* 取代為有效的 AWS 帳號。

步驟 9：執行工作流程

由於您指定工作流程是run-on-demand，因此您必須手動啟動工作流程。

- 在藍圖頁面上，選取工作流程 `lakeformationcloudtrailtest`，然後在動作功能表上，選擇開始。

  當工作流程執行時，您可以在上次執行狀態欄中檢視其進度。偶爾選擇重新整理按鈕。

  狀態會從 RUNNING 進入探索、匯入，進入 COMPLETED。

當工作流程完成時：

- Data Catalog 會有新的中繼資料表。
- 您的 CloudTrail 日誌將擷取到資料湖中。

如果工作流程失敗，請執行下列動作：

a. 選取工作流程，然後在動作功能表上，選擇檢視圖形。

   工作流程會在 AWS Glue主控台中開啟。

b. 確認已選取工作流程，然後選擇 History (歷史記錄) 標籤。

c. 在歷史記錄下，選取最近的執行，然後選擇檢視執行詳細資訊。

d. 在動態（執行時間）圖形中選取失敗的任務或爬蟲程式，並檢閱錯誤訊息。失敗的節點為紅色或黃色。

步驟 10：在資料表上授予 SELECT

您必須授予新 Data Catalog 資料表的 SELECT 許可，資料分析師才能查詢資料表指向的資料。

> **Note**
>
> 工作流程會自動將所建立資料表的SELECT許可授予執行該資料表的使用者。由於資料湖管理員執行此工作流程，您必須SELECT將 授予資料分析師。

1. 在 Lake Formation 主控台的導覽窗格中的資料目錄下，選擇資料庫。
2. 選擇`lakeformation_cloudtrail`資料庫，然後從動作下拉式清單中，選擇許可標題下的授予。
3. 在授予資料許可對話方塊中，進行下列選擇：

   a. 在主體下，針對 IAM 使用者和角色，選擇 `datalake_user`。

   b. 在 LF 標籤或目錄資源下，選擇具名資料目錄資源。

   c. 對於資料庫，應已選取`lakeformation_cloudtrail`資料庫。

   d. 針對資料表，選擇 `cloudtrailtest-cloudtrail`。

   e. 在資料表和資料欄許可下，選擇選取。

4. 選擇 Grant (授予)。

下一個步驟會以資料分析師身分執行。

步驟 11：使用 查詢資料湖 Amazon Athena

使用 Amazon Athena 主控台查詢資料湖中的 CloudTrail 資料。

1. 開啟 Athena 主控台，網址為 https://console.aws.amazon.com/athena/：//datalake_user。
2. 如有必要，請選擇入門以繼續前往 Athena 查詢編輯器。
3. 針對資料來源，選擇 AwsDataCatalog。
4. 針對 Database (資料庫)，輸入 lakeformation_cloudtrail。

   資料表清單會填入。
5. 在資料表 旁的溢位選單 （ 水平排列 3 個點 ） 上cloudtrailtest-cloudtrail，選擇預覽資料表，然後選擇執行。

   查詢會執行並顯示 10 列的資料。

   如果您之前未使用 Athena，您必須先在 Athena 主控台中設定 Amazon S3 位置，以儲存查詢結果。datalake_user 必須有必要的許可，才能存取您選擇的 Amazon S3 儲存貯體。

> **Note**
>
> 現在您已完成教學課程，請將資料許可和資料位置許可授予組織中的主體。

從 Lake Formation 中的 JDBC 來源建立資料湖

本教學課程會引導您完成在 AWS Lake Formation 主控台上執行的步驟，以使用 Lake Formation 從 JDBC 來源建立和載入您的第一個資料湖。

主題

- 目標對象
- JDBC 教學課程先決條件
- 步驟 1：建立資料分析師使用者
- 步驟 2：在 中建立連線 AWS Glue

- 步驟 3：為資料湖建立 Amazon S3 儲存貯體
- 步驟 4：註冊 Amazon S3 路徑
- 步驟 5：授予資料位置許可
- 步驟 6：在 Data Catalog 中建立資料庫
- 步驟 7：授予資料許可
- 步驟 8：使用藍圖建立工作流程
- 步驟 9：執行工作流程
- 步驟 10：在資料表上授予 SELECT
- 步驟 11：使用 查詢資料湖 Amazon Athena
- 步驟 12：使用 Amazon Redshift Spectrum 查詢資料湖中的資料
- 步驟 13：使用 Amazon Redshift Spectrum 授予或撤銷 Lake Formation 許可

目標對象

下表列出此 AWS Lake Formation JDBC 教學課程中使用的角色。

角色	描述
IAM 管理員	可建立 AWS Identity and Access Management (IAM) 使用者和角色以及 Amazon Simple Storage Service (Amazon S3) 儲存貯體的使用者。具有 `AdministratorAccess` AWS 受管政策。
資料湖管理員	可以存取 Data Catalog、建立資料庫，以及將 Lake Formation 許可授予其他使用者的使用者。IAM 許可少於 IAM 管理員，但足以管理資料湖。
資料分析	可對資料湖執行查詢的使用者。僅具有執行查詢的足夠許可。
工作流程角色	具有執行工作流程所需 IAM 政策的角色。

如需完成教學課程的先決條件資訊，請參閱 JDBC 教學課程先決條件。

JDBC 教學課程先決條件

開始 AWS Lake Formation JDBC 教學課程之前，請確定您已執行下列動作：

- 完成 Lake Formation 入門 中的任務。
- 決定您要用於教學課程的 JDBC 可存取資料存放區。
- 收集建立 JDBC 類型AWS Glue連線所需的資訊。此資料目錄物件包含資料存放區的 URL、登入憑證，以及如果資料存放區是在 Amazon Virtual Private Cloud (Amazon VPC) 中建立的，則為額外的 VPC 特定組態資訊。如需詳細資訊，請參閱《 AWS Glue 開發人員指南》中的在AWS Glue資料目錄中定義連線。

本教學課程假設您熟悉 AWS Identity and Access Management (IAM)。如需 IAM 的相關資訊，請參閱 IAM 使用者指南。

若要開始使用，請繼續 the section called “步驟 1：建立資料分析師使用者”。

步驟 1：建立資料分析師使用者

在此步驟中，您會建立 AWS Identity and Access Management (IAM) 使用者，以成為資料湖所在的資料分析師 AWS Lake Formation。

此使用者具有查詢資料湖的最小許可集。

1. 在 https://console.aws.amazon.com/iam 開啟 IAM 主控台。以您在 中建立的管理員使用者身分登入，建立具有管理存取權的使用者或以 `AdministratorAccess` AWS 受管政策的使用者身分登入。
2. `datalake_user` 使用下列設定建立名為 的使用者：
   - 啟用 AWS Management Console 存取。
   - 設定密碼，不需要重設密碼。
   - 連接 `AmazonAthenaFullAccess` AWS 受管政策。
   - 連接下列內嵌政策。將政策命名為 `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": "*"
        }
    ]
}
```

步驟 2：在 中建立連線 AWS Glue

> **Note**
>
> 如果您已有與 JDBC 資料來源的AWS Glue連線，請略過此步驟。

AWS Lake Formation 透過 AWS Glue連線存取 JDBC 資料來源。連線是 Data Catalog 物件，其中包含連線至資料來源所需的所有資訊。您可以使用 AWS Glue主控台建立連線。

建立連線

1. 在 AWS Glue開啟 主控台https://console.aws.amazon.com/glue/，並以您在 中建立的管理員使用者身分登入建立具有管理存取權的使用者。
2. 在導覽窗格中，於 Data catalog (Data Catalog) 下選擇 Connections (連線)。
3. 在連接器頁面上，選擇建立連線。
4. 在選擇資料來源頁面上，選擇 JDBC 做為連線類型。然後選擇下一步。
5. 繼續執行連線精靈並儲存連線。

   如需建立連線的資訊，請參閱《 AWS Glue 開發人員指南》中的 AWS Glue JDBC 連線屬性。

步驟 3：為資料湖建立 Amazon S3 儲存貯體

在此步驟中，您會建立 Amazon Simple Storage Service (Amazon S3) 儲存貯體，該儲存貯體將成為資料湖的根位置。

1. 在 https://console.aws.amazon.com/s3/ 開啟 Amazon S3 主控台，並以您在 中建立的管理員使用者身分登入建立具有管理存取權的使用者。
2. 選擇建立儲存貯體，然後瀏覽精靈以建立名為 的儲存貯體`<yourName>-datalake-tutorial`，其中 `<yourName>` 是您的第一個初始名稱和姓氏。例如：`jdoe-datalake-tutorial`。

   如需建立 Amazon S3 儲存貯體的詳細說明，請參閱《Amazon Simple Storage Service 使用者指南》中的如何建立 S3 儲存貯體？。

步驟 4：註冊 Amazon S3 路徑

在此步驟中，您將 Amazon Simple Storage Service (Amazon S3) 路徑註冊為資料湖的根位置。

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。以資料湖管理員身分登入。
2. 在導覽窗格的管理下，選擇資料湖位置。
3. 選擇註冊位置，然後選擇瀏覽。
4. 選取您先前建立的儲存`<yourName>-datalake-tutorial`貯體，接受預設 IAM 角色 `AWSServiceRoleForLakeFormationDataAccess`，然後選擇註冊位置。

   如需註冊位置的詳細資訊，請參閱 將 Amazon S3 位置新增至您的資料湖。

步驟 5：授予資料位置許可

主體必須在資料湖位置上具有資料位置許可，才能建立指向該位置的資料目錄資料表或資料庫。您必須將資料位置許可授予工作流程的 IAM 角色，工作流程才能寫入資料擷取目的地。

1. 在 Lake Formation 主控台的導覽窗格中的許可下，選擇資料位置。
2. 選擇授予，然後在授予許可對話方塊中，執行下列動作：

   a. 針對 IAM 使用者和角色，選擇 `LakeFormationWorkflowRole`。

   b. 針對儲存位置，選擇您的儲存`<yourName>-datalake-tutorial`貯體。

3. 選擇 Grant (授予)。

如需資料位置許可的詳細資訊，請參閱 Underlying data access control。

步驟 6：在 Data Catalog 中建立資料庫

Lake Formation Data Catalog 中的中繼資料表會存放在資料庫中。

1. 在 Lake Formation 主控台的導覽窗格中的資料目錄下，選擇資料庫。
2. 選擇建立資料庫，然後在資料庫詳細資訊下，輸入名稱 `lakeformation_tutorial`。
3. 將其他欄位保留空白，然後選擇建立資料庫。

步驟 7：授予資料許可

您必須授予在 Data Catalog 中建立中繼資料表的許可。由於工作流程使用角色 執行`LakeFormationWorkflowRole`，您必須將這些許可授予角色。

1. 在 Lake Formation 主控台的導覽窗格中的許可下，選擇資料湖許可。
2. 選擇授予，然後在授予資料許可對話方塊中，執行下列動作：
   a. 在主體下，針對 IAM 使用者和角色，選擇 `LakeFormationWorkflowRole`。
   b. 在 LF 標籤或目錄資源下，選擇具名資料目錄資源。
   c. 針對資料庫，選擇您先前建立的資料庫 `lakeformation_tutorial`。
   d. 在資料庫許可下，選取建立資料表、修改和捨棄，並在選取時清除超級。
3. 選擇 Grant (授予)。

如需授予 Lake Formation 許可的詳細資訊，請參閱 Lake Formation 許可概觀 。

步驟 8：使用藍圖建立工作流程

AWS Lake Formation 工作流程會產生AWS Glue任務、爬蟲程式和觸發程序，以探索資料並將其擷取至您的資料湖。您可以根據其中一個預先定義的 Lake Formation 藍圖建立工作流程。

1. 在 Lake Formation 主控台的導覽窗格中，選擇藍圖，然後選擇使用藍圖。
2. 在使用藍圖頁面的藍圖類型下，選擇資料庫快照。

3. 在匯入來源下，針對資料庫連線，選擇您剛建立的連線、 `datalake-tutorial`，或為您的資料來源選擇現有的連線。
4. 對於來源資料路徑，輸入要從中擷取資料的路徑，格式為 *<database>*/*<schema>*/*<table>*。

   您可以用百分比 (%) 萬用字元取代結構描述或資料表。對於支援結構描述的資料庫，請輸入 *<database>*/*<schema>*/%，以符合 *<database>* 內 *<schema>* 中的所有資料表。Oracle 資料庫和 MySQL 不支援路徑中的結構描述；而是輸入 *<database>*/%。對於 Oracle 資料庫，*<database>* 是系統識別符 (SID)。

   例如，如果 Oracle 資料庫具有 `orcl`做為其 SID，請輸入 `orcl/%` 以符合 JDCB 連線中指定的使用者可存取的所有資料表。

   > **Important**
   >
   > 此欄位會區分大小寫。

5. 在匯入目標下，指定下列參數：

目標資料庫	`lakeformation_tutorial`
目標儲存位置	`s3://`*<yourName>* `-datalake-tutorial`
資料格式	（選擇 Parquet 或 CSV)

6. 針對匯入頻率，選擇隨需執行。
7. 在匯入選項下，指定下列參數：

工作流程名稱	`lakeformationjdbctest`
IAM 角色	`LakeFormationWorkflowRole`
資料表字首	`jdbctest` **Note** 必須是小寫。

8. 選擇建立，然後等待主控台報告工作流程已成功建立。

> **Tip**
>
> 您是否收到下列錯誤訊息？
> `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...`
> 若是如此，請檢查您是否已將資料湖管理員使用者的內嵌政策中的 *<account-id>* 取代為有效的 AWS 帳號。

步驟 9：執行工作流程

由於您指定工作流程是run-on-demand，因此您必須在其中手動啟動工作流程 AWS Lake Formation。

1. 在 Lake Formation 主控台的藍圖頁面上，選取工作流程 `lakeformationjdbctest`。
2. 選擇動作，然後選擇開始。
3. 當工作流程執行時，請在上次執行狀態欄中檢視其進度。偶爾選擇重新整理按鈕。

   狀態會從 RUNNING 到 Discovering、匯入，再到 COMPLETED。

   當工作流程完成時：

   - Data Catalog 有新的中繼資料表。
   - 您的資料會擷取到資料湖中。

   如果工作流程失敗，請執行下列動作：

   a. 選取工作流程。選擇動作，然後選擇檢視圖形。

      工作流程會在 AWS Glue主控台中開啟。

   b. 選取工作流程，然後選擇歷史記錄索引標籤。

   c. 選取最近的執行，然後選擇檢視執行詳細資訊。

   d. 在動態（執行時間）圖形中選取失敗的任務或爬蟲程式，然後檢閱錯誤訊息。失敗的節點為紅色或黃色。

步驟 10：在資料表上授予 SELECT

您必須授予 中新 Data Catalog 資料表的 SELECT 許可， AWS Lake Formation 資料分析師才能查詢資料表指向的資料。

> **Note**
>
> 工作流程會自動將所建立資料表的SELECT許可授予執行該資料表的使用者。由於資料湖管理員執行此工作流程，您必須SELECT將 授予資料分析師。

1. 在 Lake Formation 主控台的導覽窗格中的許可下，選擇資料湖許可。
2. 選擇授予，然後在授予資料許可對話方塊中，執行下列動作：
   a. 在主體下，針對 IAM 使用者和角色，選擇 `datalake_user`。
   b. 在 LF 標籤或目錄資源下，選擇具名資料目錄資源。
   c. 針對資料庫，選擇 `lakeformation_tutorial`。

      資料表清單即會填入。
   d. 對於資料表，請從資料來源中選擇一或多個資料表。
   e. 在資料表和資料欄許可下，選擇選取。
3. 選擇 Grant (授予)。

下一個步驟會以資料分析師身分執行。

步驟 11：使用 查詢資料湖 Amazon Athena

使用 Amazon Athena 主控台查詢資料湖中的資料。

1. 在 https：//https://console.aws.amazon.com/athena/ 開啟 Athena 主控台，並以資料分析師使用者 身分登入`datalake_user`。
2. 如有必要，請選擇開始使用以繼續前往 Athena 查詢編輯器。
3. 針對資料來源，選擇 AwsDataCatalog。
4. 針對 Database (資料庫)，輸入 `lakeformation_tutorial`。

   資料表清單即會填入。

5. 在其中一個資料表旁的快顯功能表中，選擇預覽資料表。

   查詢會執行並顯示 10 列的資料。

步驟 12：使用 Amazon Redshift Spectrum 查詢資料湖中的資料

您可以設定 Amazon Redshift Spectrum 來查詢匯入 Amazon Simple Storage Service (Amazon S3) 資料湖的資料。首先，建立用於啟動 Amazon Redshift 叢集和查詢 Amazon S3 資料的 AWS Identity and Access Management (IAM) 角色。然後，將您要查詢之資料表的`Select`許可授予此角色。然後，授予使用者使用 Amazon Redshift 查詢編輯器的許可。最後，建立 Amazon Redshift 叢集並執行查詢。

您以管理員身分建立叢集，並以資料分析師身分查詢叢集。

如需 Amazon Redshift Spectrum 的詳細資訊，請參閱《Amazon Redshift 資料庫開發人員指南》中的使用 Amazon Redshift Spectrum 查詢外部資料。

設定執行 Amazon Redshift 查詢的許可

1. 前往 https://console.aws.amazon.com/iam/ 開啟 IAM 主控台。以您在 中建立的管理員使用者身分登入 建立具有管理存取權的使用者（使用者名稱 `Administrator`），或以具有 `AdministratorAccess` AWS 受管政策的使用者身分登入。
2. 在導覽窗格中，選擇政策。

   如果這是您第一次選擇 Policies (政策)，將會顯示 Welcome to Managed Policies (歡迎使用受管政策) 頁面。選擇 Get Started (開始使用)。
3. 選擇建立政策。
4. 請選擇 JSON 標籤。
5. 貼上下列 JSON 政策文件。

```
{
    "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. 完成時，選擇 Review (檢閱) 以檢閱該政策。政策驗證程式會回報任何語法錯誤。
7. 在檢閱政策頁面上，輸入您所建立**RedshiftLakeFormationPolicy**政策的名稱。輸入 Description (說明) (選用)。檢閱政策 Summary (摘要) 來查看您的政策所授予的許可。然後選擇 Create policy (建立政策) 來儲存您的工作。
8. 在 IAM 主控台的導覽窗格中，選擇 Roles (角色)，然後選擇 Create role (建立角色)。
9. 對於 Select trusted entity (選取信任的實體) 區段，選擇 AWS service (AWS 服務)。
10. 選擇 Amazon Redshift 服務以擔任此角色。
11. 選擇服務的 Redshift Customizable (Redshift 可自訂) 使用案例。然後選擇下一步：許可。
12. 搜尋您建立的許可政策 `RedshiftLakeFormationPolicy`，然後選取清單中政策名稱旁的核取方塊。
13. 選擇下一步：標籤。
14. 選擇下一步：檢閱。
15. 在 Role name (角色名稱) 中，輸入名稱 **RedshiftLakeFormationRole**。
16. (選用) 在 Role description (角色說明) 中，輸入新角色的說明。
17. 檢閱角色，然後選擇 Create role (建立角色)。

授予要在 Lake Formation 資料庫中查詢之資料表的**Select**許可

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。以資料湖管理員身分登入。
2. 在導覽窗格的許可下，選擇資料湖許可，然後選擇授予。
3. 請提供下列資訊：

- 針對 IAM 使用者和角色，選擇您建立的 IAM 角色 RedshiftLakeFormationRole。當您執行 Amazon Redshift 查詢編輯器時，其會使用此 IAM 角色來取得資料的許可。
- 針對 Database (資料庫)，輸入 lakeformation_tutorial。

  表格清單會填入 。
- 在資料表中，選擇要查詢的資料來源中的資料表。
- 選擇選取資料表許可。

4. 選擇 Grant (授予)。

設定 Amazon Redshift Spectrum 並執行查詢

1. 在 開啟 Amazon Redshift 主控台https://console.aws.amazon.com/redshift。以使用者 身分登入Administrator。
2. 選擇 建立叢集。
3. 在建立叢集頁面上，redshift-lakeformation-demo為叢集識別符輸入 。
4. 針對節點類型，選取 dc2.large。
5. 向下捲動，然後在資料庫組態下，輸入或接受這些參數：
   - 管理員使用者名稱： awsuser
   - 管理員使用者密碼：(*Choose a password*)
6. 展開叢集許可，對於可用的 IAM 角色，選擇 RedshiftLakeFormationRole。接著選擇 Add IAM role (新增 IAM 角色)。
7. 如果您必須使用與預設值 5439 不同的連接埠，請在其他組態旁關閉使用預設值選項。展開資料庫組態的 區段，然後輸入新的資料庫連接埠號碼。
8. 選擇 建立叢集。

   叢集頁面會載入。
9. 等到叢集狀態變成可用。定期選擇重新整理圖示。
10. 授予資料分析師對叢集執行查詢的許可。若要這樣做，請完成下列步驟。
    a. 在 https://console.aws.amazon.com/iam/：// 開啟 IAM 主控台，並以Administrator使用者身分登入。
    b. 在導覽窗格中，選擇使用者，然後將下列受管政策連接至使用者 datalake_user。

- AmazonRedshiftQueryEditor
- AmazonRedshiftReadOnlyAccess

11. 登出 Amazon Redshift 主控台，並以使用者 身分重新登入datalake_user。
12. 在左側垂直工具列中，選擇 EDITOR 圖示以開啟查詢編輯器並連線至叢集。如果出現連線至資料庫對話方塊，請選擇叢集名稱 redshift-lakeformation-demo，然後輸入資料庫名稱 **dev**、使用者名稱 **awsuser**和您建立的密碼。選擇 Connect to database (連線至資料庫)。

> **Note**
>
> 如果未提示您輸入連線參數，且已在查詢編輯器中選取另一個叢集，請選擇變更連線以開啟連線至資料庫對話方塊。

13. 在新查詢 1 文字方塊中，輸入並執行下列陳述式，將 Lake Formation lakeformation_tutorial中的資料庫映射至 Amazon Redshift 結構描述名稱 redshift_jdbc：

> **Important**
>
> 將 *<account-id>* 取代為有效的 AWS 帳號，並將 *<region>* 取代為有效的 AWS 區域名稱（例如 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. 在選取結構描述下的結構描述清單中，選擇 redshift_jdbc。

    表格清單會填入 。查詢編輯器只會顯示您獲得 Lake Formation 資料湖許可的資料表。

15. 在資料表名稱旁的快顯功能表中，選擇預覽資料。

    Amazon Redshift 會傳回前 10 列。

    您現在可以針對您擁有許可的資料表和資料欄執行查詢。

步驟 13：使用 Amazon Redshift Spectrum 授予或撤銷 Lake Formation 許可

Amazon Redshift 支援使用修改後的 SQL 陳述式授予和撤銷資料庫和資料表的 Lake Formation 許可。這些陳述式類似於現有的 Amazon Redshift 陳述式。如需詳細資訊，請參閱《Amazon Redshift 資料庫開發人員指南》中的 GRANT 和 REVOKE。

在 Lake Formation 中設定開放資料表儲存格式的許可

AWS Lake Formation 支援管理開放資料表格式 (OTFs的存取許可，例如 Apache Iceberg、Apache Hudi 和 Linux 基礎 Delta Lake。在本教學課程中，您將了解如何 AWS Glue Data Catalog 使用 建立 Iceberg、Hudi 和 Delta Lake 搭配 symlink 資訊清單表格 AWS Glue、使用 Lake Formation 設定精細許可，以及使用 Amazon Athena 查詢資料。

> **Note**
>
> AWS 分析服務不支援所有交易資料表格式。如需詳細資訊，請參閱使用其他 AWS 服務。本教學課程僅涵蓋使用 AWS Glue 任務在 Data Catalog 中手動建立新的資料庫和資料表。

本教學課程包含快速設定的 AWS CloudFormation 範本。您可以檢閱並自訂它，以符合您的需求。

主題

- 目標對象
- 先決條件
- 步驟 1：佈建您的 資源
- 步驟 2：設定 Iceberg 資料表的許可
- 步驟 3：設定 Hudi 資料表的許可
- 步驟 4：設定 Delta Lake 資料表的許可
- 步驟 5：清除 AWS 資源

目標對象

本教學課程適用於 IAM 管理員、資料湖管理員和商業分析師。下表列出本教學課程中用於使用 Lake Formation 建立受管資料表的角色。

角色	描述
IAM 管理員	可建立 IAM 使用者和角色以及 Amazon S3 儲存貯體的使用者。具有 `AdministratorAccess` AWS 受管政策。
資料湖管理員	可以存取 Data Catalog、建立資料庫，以及將 Lake Formation 許可授予其他使用者的使用者。IAM 許可少於 IAM 管理員，但足以管理資料湖。
業務分析師	可對資料湖執行查詢的使用者。具有執行查詢的許可。

先決條件

開始本教學課程之前，您必須擁有 AWS 帳戶 可以具有正確許可的使用者身分登入的 。如需詳細資訊，請參閱 註冊 AWS 帳戶 和 建立具有管理存取權的使用者。

本教學課程假設您熟悉 IAM 角色和政策。如需 IAM 的相關資訊，請參閱 IAM 使用者指南。

您需要設定下列 AWS 資源才能完成本教學課程：

- 資料湖管理員使用者
- Lake Formation 資料湖設定
- Amazon Athena 引擎版本 3

建立資料湖管理員

1. 以管理員使用者身分登入位於 https://console.aws.amazon.com/lakeformation/ 的 Lake Formation 主控台。您將在美國東部（維吉尼亞北部）區域建立本教學課程的資源。
2. 在 Lake Formation 主控台的導覽窗格中，於許可下，選擇管理角色和任務。
3. 選取資料湖管理員下的選擇管理員。
4. 在快顯視窗中，管理資料湖管理員，在 IAM 使用者和角色下，選擇 IAM 管理員使用者。
5. 選擇 Save (儲存)。

啟用資料湖設定

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。在導覽窗格中的資料目錄下，選擇設定。取消勾選下列項目：
   - 僅對新資料庫使用 IAM 存取控制。
   - 僅對新資料庫中的新資料表使用 IAM 存取控制。
2. 在跨帳戶版本設定下，選擇第 3 版作為跨帳戶版本。
3. 選擇 Save (儲存)。

將 Amazon Athena 引擎升級至第 3 版

1. 開啟 Athena 主控台，網址為 https：//https://console.aws.amazon.com/athena/。
2. 選取工作群組，然後選取主要工作群組。
3. 確定工作群組的最小版本為 3。如果不是，請編輯工作群組，選擇手動升級查詢引擎，然後選取第 3 版。
4. 選擇 Save changes (儲存變更)。

步驟 1：佈建您的 資源

本節說明如何使用 AWS CloudFormation 範本設定 AWS 資源。

使用 AWS CloudFormation 範本建立資源

1. 以美國東部（維吉尼亞北部）區域中的 IAM 管理員身分登入 AWS CloudFormation 主控台，網址為 https://console.aws.amazon.com/cloudformation：//www.healthnet.com。
2. 選擇啟動堆疊。
3. 在建立堆疊畫面上選擇下一步。
4. 輸入堆疊名稱。
5. 選擇 Next (下一步)。
6. 在下一頁中，選擇下一步。
7. 檢閱最終頁面上的詳細資訊，然後選取我確認 AWS CloudFormation 可能會建立 IAM 資源。
8. 選擇 Create (建立)。

   堆疊建立最多可能需要兩分鐘。

啟動雲端形成堆疊會建立下列資源：

- lf-otf-datalake-123456789012 – 存放資料的 Amazon S3 儲存貯體

> **Note**
>
> 附加至 Amazon S3 儲存貯體名稱的帳戶 ID 會取代為您的帳戶 ID。

- lf-otf-tutorial-123456789012 – Amazon S3 儲存貯體，用於儲存查詢結果和 AWS Glue 任務指令碼
- lficebergdb – AWS Glue Iceberg 資料庫
- lfhudidb – AWS Glue Hudi 資料庫
- lfdeltadb – AWS Glue Delta 資料庫
- native-iceberg-create – 在 Data Catalog 中建立 Iceberg 資料表 AWS Glue 的任務
- native-hudi-create – 在 Data Catalog 中建立 Hudi 資料表 AWS Glue 的任務
- native-delta-create – AWS Glue job，可在 Data Catalog 中建立 Delta 資料表
- LF-OTF-GlueServiceRole – 您傳遞給 AWS Glue 以執行任務的 IAM 角色。此角色已連接必要政策，以存取資料目錄、Amazon S3 儲存貯體等資源。
- LF-OTF-RegisterRole – 向 Lake Formation 註冊 Amazon S3 位置的 IAM 角色。此角色已`LF-Data-Lake-Storage-Policy`連接至角色。
- lf-consumer-analystuser – 使用 Athena 查詢資料的 IAM 使用者
- lf-consumer-analystuser-credentials – 存放在 的資料分析師使用者密碼 AWS Secrets Manager

堆疊建立完成後，導覽至輸出索引標籤並記下下列項目的值：

- AthenaQueryResultLocation – Athena 查詢輸出的 Amazon S3 位置
- BusinessAnalystUserCredentials – 資料分析師使用者的密碼

  若要擷取密碼值：

  1. 導覽至 Secrets Manager 主控台以選擇`lf-consumer-analystuser-credentials`值。
  2. 在 Secret value (秘密值) 區段，選擇 Retrieve secret value (擷取秘密值)。
  3. 記下密碼的秘密值。

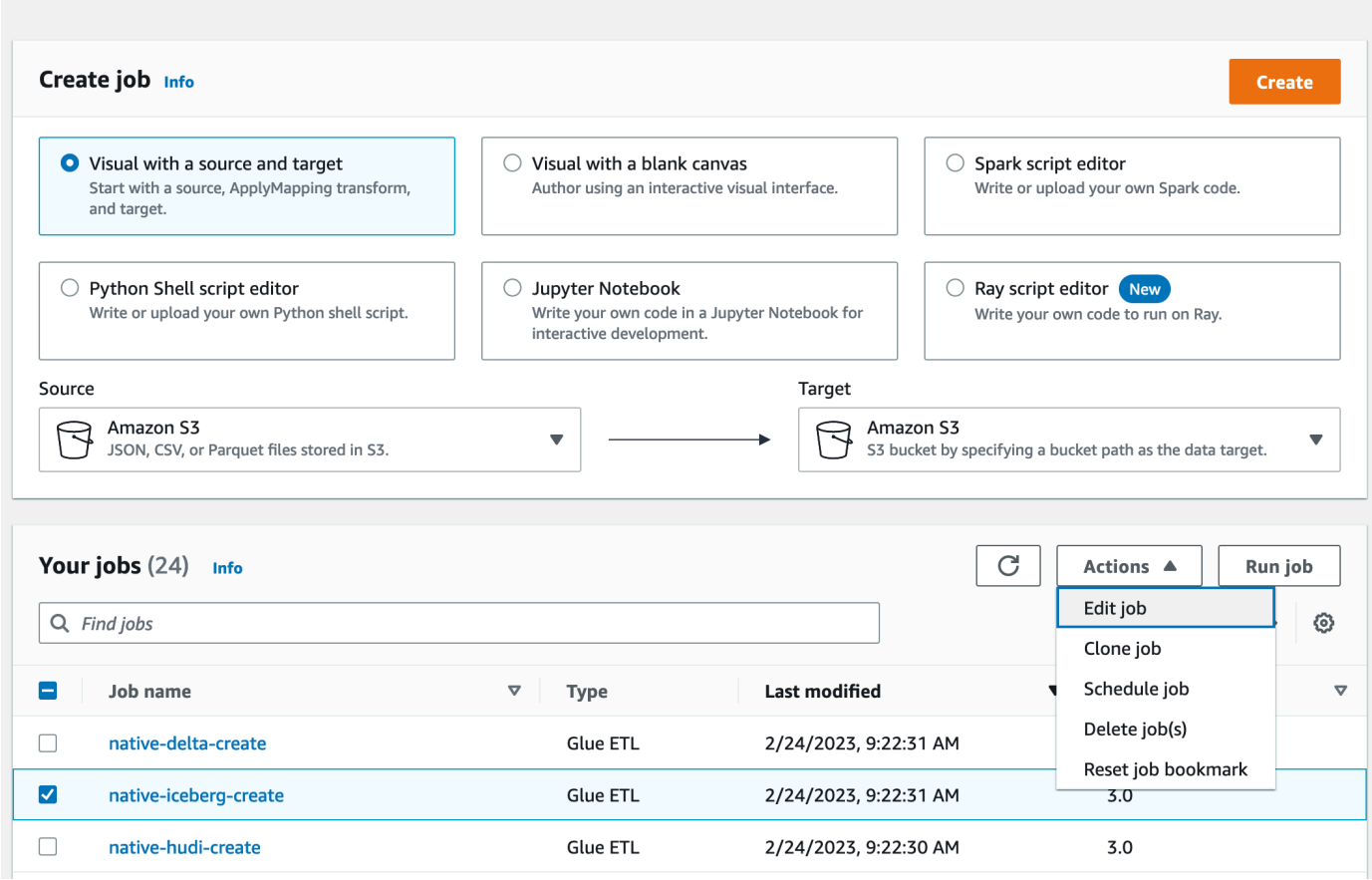
步驟 2：設定 Iceberg 資料表的許可

在本節中，您將了解如何在 中建立 Iceberg 資料表 AWS Glue Data Catalog、在 中設定資料許可 AWS Lake Formation，以及使用 Amazon Athena 查詢資料。

建立 Iceberg 資料表

在此步驟中，您將執行在 Data Catalog 中建立 Iceberg 交易資料表 AWS Glue 的任務。

1. 在美國東部（維吉尼亞北部）區域中，以資料湖管理員使用者身分開啟 AWS Glue 主控台，網址為 https://console.aws.amazon.com/glue/：//www.healthnet.com。
2. 從左側導覽窗格中選擇任務。
3. 選取 `native-iceberg-create`。



4. 在動作下，選擇編輯任務。
5. 在任務詳細資訊下，展開進階屬性，並勾選使用 AWS Glue Data Catalog 做為 Hive 中繼存放區，以在 中新增資料表中繼資料旁的方塊 AWS Glue Data Catalog。這會指定 AWS Glue Data Catalog 做為任務中所用 Data Catalog 資源的中繼存放區，並啟用 Lake Formation 許可，以便稍後套用至目錄資源。

6. 選擇 Save (儲存)。
7. 選擇執行。您可以在任務執行時檢視任務的狀態。

   如需 AWS Glue 任務的詳細資訊，請參閱《 AWS Glue 開發人員指南》中的在 AWS Glue 主控台上使用任務。

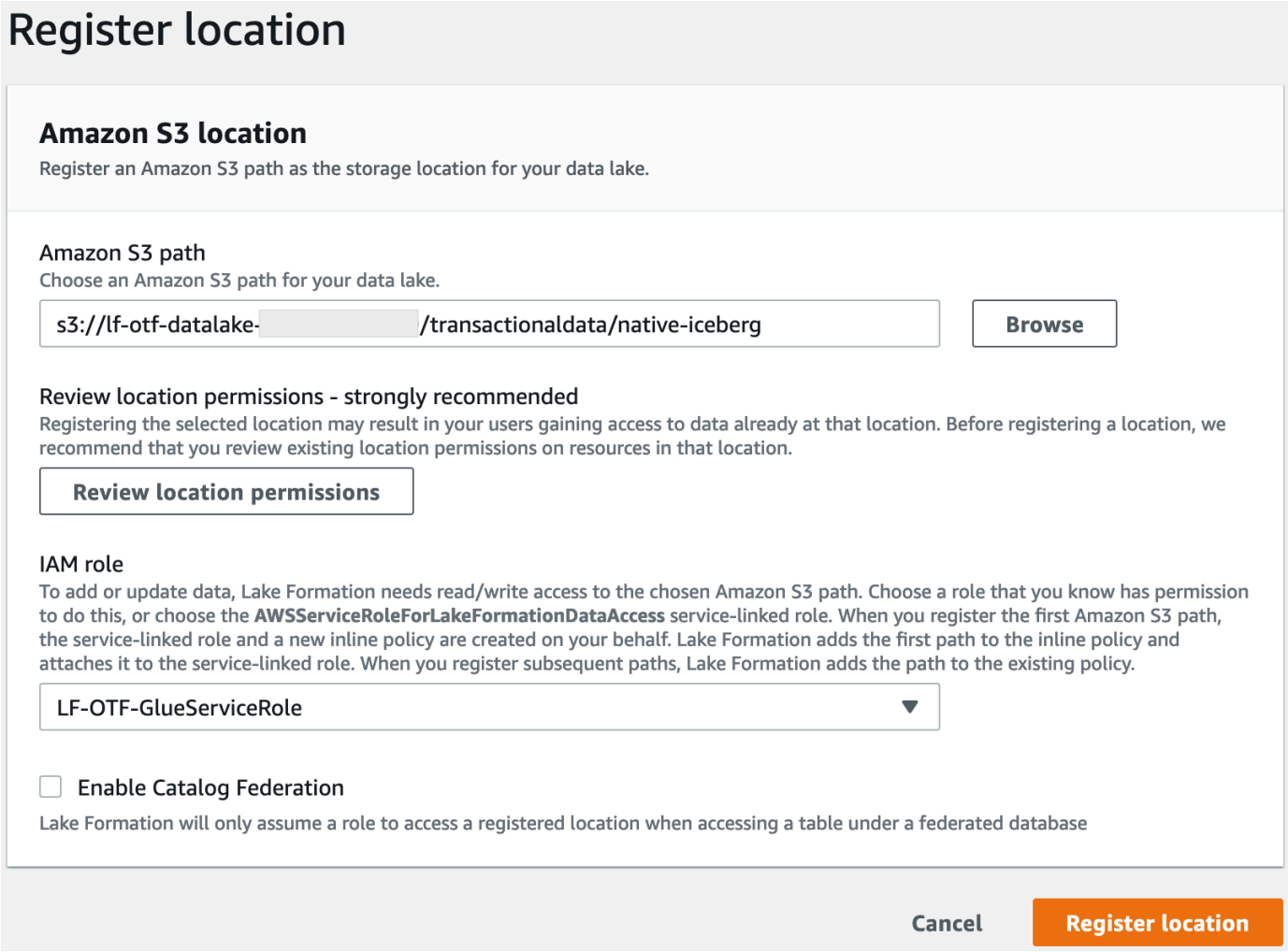
   此任務會在`lficebergdb`資料庫中建立名為 `product`的 Iceberg 資料表。在 Lake Formation 主控台中驗證產品資料表。

向 Lake Formation 註冊資料位置

接著，將 Amazon S3 路徑註冊為資料湖的位置。

1. 以資料湖管理員使用者身分開啟位於 https://console.aws.amazon.com/lakeformation/：// 的 Lake Formation 主控台。
2. 在導覽窗格的註冊和擷取下，選擇資料位置。
3. 在主控台的右上角，選擇註冊位置。
4. 在註冊位置頁面上，輸入下列內容：

   - Amazon S3 路徑 – 選擇瀏覽並選取 `lf-otf-datalake-123456789012`。按一下 Amazon S3 根位置旁的向右箭頭 (>) 以導覽至`s3/buckets/lf-otf-datalake-123456789012/transactionaldata/native-iceberg`位置。
   - IAM 角色 – 選擇 `LF-OTF-RegisterRole`作為 IAM 角色。
   - 選擇註冊位置。

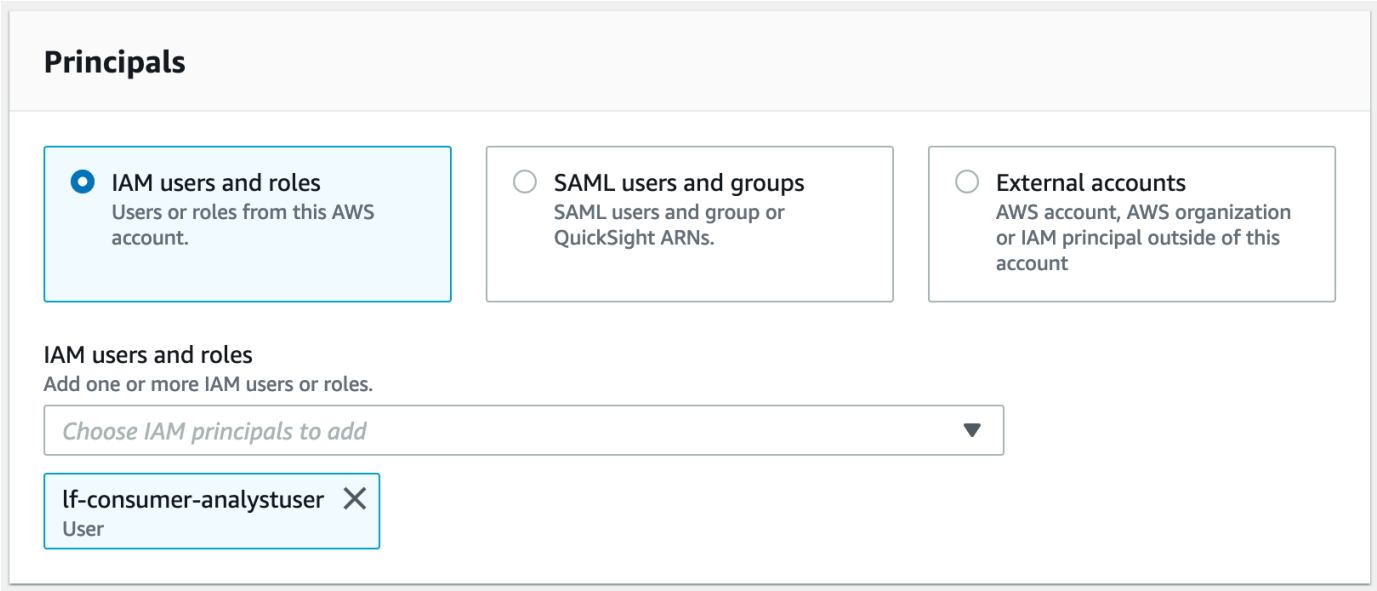


如需向 Lake Formation 註冊資料位置的詳細資訊，請參閱 [將 Amazon S3 位置新增至您的資料湖](#)。

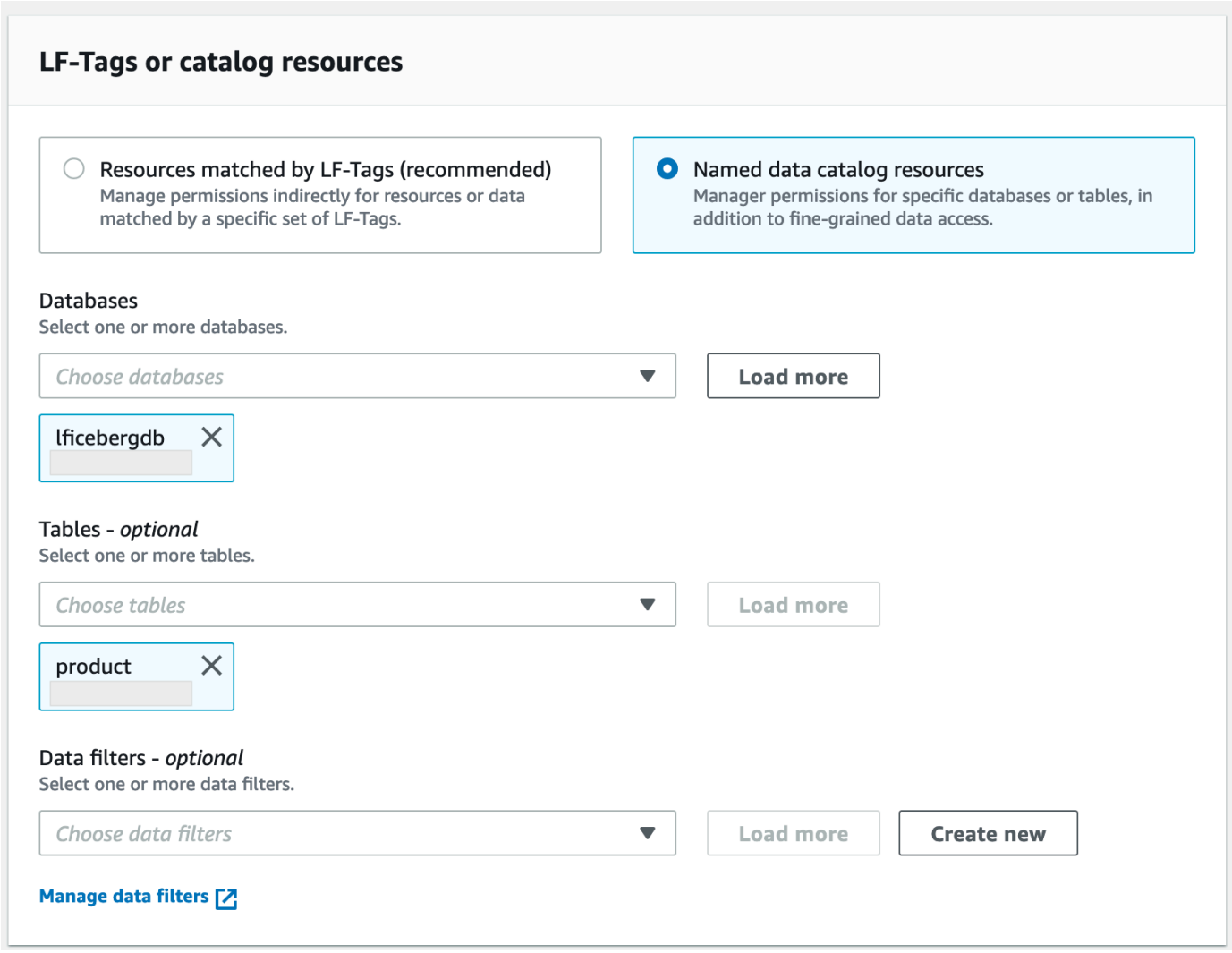
在 Iceberg 資料表上授予 Lake Formation 許可

在此步驟中，我們會將資料湖許可授予商業分析師使用者。

1. 在資料湖許可下，選擇授予。
2. 在授予資料許可畫面上，選擇 IAM 使用者和角色。
3. `lf-consumer-analystuser` 從下拉式清單中選擇。



4. 選擇具名資料目錄資源。
5. 針對資料庫，選擇 `lficebergdb`。
6. 針對資料表，選擇 `product`。



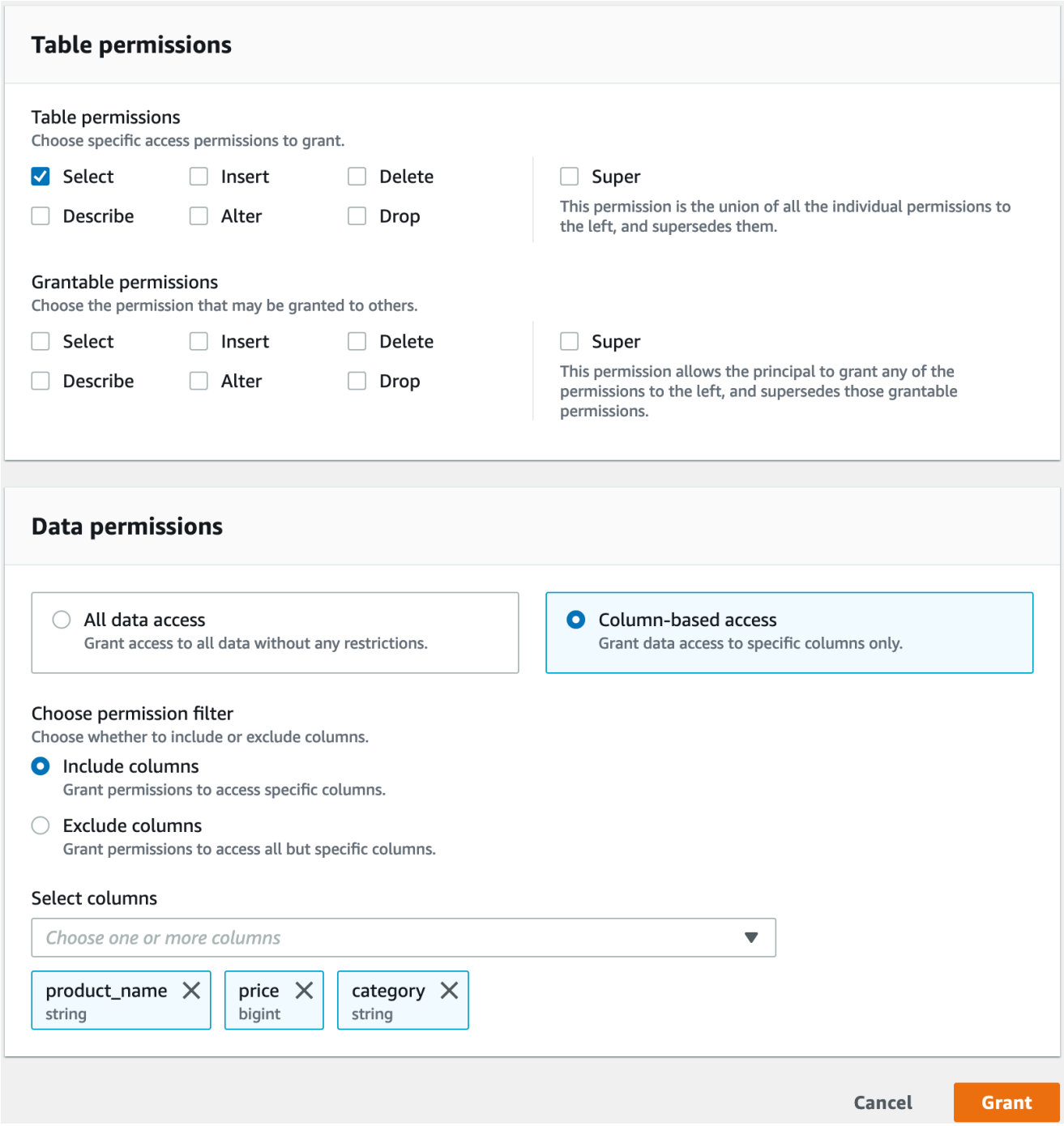
7. 接下來，您可以指定資料欄來授予資料欄型存取。

   a. 在資料表許可下，選擇選取。

   b. 在資料許可下，選擇資料欄型存取，選擇包含資料欄。

   c. 選擇 `product_name`、 `price`和 `category`欄。

   d. 選擇 Grant (授予)。



使用 Athena 查詢 Iceberg 資料表

現在您可以開始查詢您使用 Athena 建立的 Iceberg 資料表。如果您是第一次在 Athena 中執行查詢，則需要設定查詢結果位置。如需詳細資訊，請參閱指定查詢結果位置。

1. 以資料湖管理員使用者身分登出，並使用先前從 AWS CloudFormation 輸出記下的密碼`lf-consumer-analystuser`，以美國東部（維吉尼亞北部）區域中的 身分登入。

2. 前往 https://console.aws.amazon.com/athena/ 開啟 Athena 主控台。
3. 選擇設定，然後選取管理。
4. 在查詢結果的位置方塊中，輸入您在 AWS CloudFormation 輸出中建立的儲存貯體路徑。複製 `AthenaQueryResultLocation`(s3：//lf-otf-tutorial-123456789012/athena-results/) 的值，然後選擇儲存。
5. 執行下列查詢來預覽存放在 Iceberg 資料表中的 10 筆記錄：

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

   如需使用 Athena 查詢 Iceberg 資料表的詳細資訊，請參閱《Amazon Athena 使用者指南》中的查詢 Iceberg 資料表。

步驟 3：設定 Hudi 資料表的許可

在本節中，您將了解如何在 中建立 Hudi 資料表 AWS Glue Data Catalog、在 中設定資料許可 AWS Lake Formation，以及使用 Amazon Athena 查詢資料。

建立 Hudi 資料表

在此步驟中，您將執行在 Data Catalog 中建立 Hudi 交易資料表 AWS Glue 的任務。

1. 在美國東部（維吉尼亞北部）區域，登入AWS Glue位於 https://console.aws.amazon.com/glue/：//www./www. https：//www.healthnet.com

   做為資料湖管理員使用者。
2. 從左側導覽窗格中選擇任務。
3. 選取 `native-hudi-create`。
4. 在動作下，選擇編輯任務。
5. 在任務詳細資訊下，展開進階屬性，並勾選使用 AWS Glue Data Catalog 做為 Hive 中繼存放區，以在 中新增資料表中繼資料旁的方塊 AWS Glue Data Catalog。這會指定 AWS Glue Data Catalog 做為任務中所用 Data Catalog 資源的中繼存放區，並啟用 Lake Formation 許可，以便稍後套用至目錄資源。
6. 選擇 Save (儲存)。
7. 選擇執行。您可以在任務執行時檢視任務的狀態。

如需 AWS Glue 任務的詳細資訊，請參閱《 AWS Glue 開發人員指南》中的在 AWS Glue 主控台上使用任務。

此任務會在 database：lfhudidb 中建立 Hudi(cow) 資料表。在 Lake Formation 主控台中驗證product資料表。

向 Lake Formation 註冊資料位置

接著，將 Amazon S3 路徑註冊為資料湖的根位置。

1. 以資料湖管理員使用者身分登入 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/：//www.healthnet.com。
2. 在導覽窗格的註冊和擷取下，選擇資料位置。
3. 在主控台的右上角，選擇註冊位置。
4. 在註冊位置頁面上，輸入下列內容：
   - Amazon S3 路徑 – 選擇瀏覽並選取 lf-otf-datalake-123456789012。按一下 Amazon S3 根位置旁的向右箭頭 (>) 以導覽至s3/buckets/lf-otf-datalake-123456789012/transactionaldata/native-hudi位置。
   - IAM 角色 – 選擇 LF-OTF-RegisterRole作為 IAM 角色。
   - 選擇註冊位置。

在 Hudi 資料表上授予資料湖許可

在此步驟中，我們會將資料湖許可授予商業分析師使用者。

1. 在資料湖許可下，選擇授予。
2. 在授予資料許可畫面上，選擇 IAM 使用者和角色。
3. lf-consumer-analystuser 從下拉式清單。
4. 選擇具名資料目錄資源。
5. 針對資料庫，選擇 lfhudidb。
6. 針對資料表，選擇 product。
7. 接下來，您可以指定資料欄來授予資料欄型存取。
   a. 在資料表許可下，選擇選取。

b. 在資料許可下，選擇資料欄型存取，選擇包含資料欄。

c. 選擇 `product_name`、 `price`和 `category`欄。

d. 選擇 Grant (授予)。

使用 Athena 查詢 Hudi 資料表

現在開始查詢您使用 Athena 建立的 Hudi 資料表。如果您是第一次在 Athena 中執行查詢，則需要設定查詢結果位置。如需詳細資訊，請參閱指定查詢結果位置。

1. 以資料湖管理員使用者身分登出，並使用先前從 AWS CloudFormation 輸出記下的密碼`lf-consumer-analystuser`，以美國東部（維吉尼亞北部）區域中的 身分登入。
2. 前往 https://console.aws.amazon.com/athena/ 開啟 Athena 主控台。
3. 選擇設定，然後選取管理。
4. 在查詢結果的位置方塊中，輸入您在 AWS CloudFormation 輸出中建立的儲存貯體路徑。複製`AthenaQueryResultLocation`(s3：//lf-otf-tutorial-123456789012/athena-results/) 和儲存的值。
5. 執行下列查詢來預覽存放在 Hudi 資料表中的 10 筆記錄：

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

如需查詢 Hudi 資料表的詳細資訊，請參閱《Amazon Athena 使用者指南》中的查詢 Hudi 資料表一節。

步驟 4：設定 Delta Lake 資料表的許可

在本節中，您將了解如何在 中建立具有符號連結資訊清單檔案的 Delta Lake 資料表 AWS Glue Data Catalog、在 中設定資料許可， AWS Lake Formation 以及使用 Amazon Athena 查詢資料。

建立 Delta Lake 資料表

在此步驟中，您將執行在 Data Catalog 中建立 Delta Lake 交易資料表 AWS Glue 的任務。

1. 在美國東部（維吉尼亞北部）區域，登入位於 https://console.aws.amazon.com/glue/：//www./ 的 AWS Glue 主控台

   做為資料湖管理員使用者。

2. 從左側導覽窗格中選擇任務。
3. 選取 native-delta-create。
4. 在動作下，選擇編輯任務。
5. 在任務詳細資訊下，展開進階屬性，並勾選使用 AWS Glue Data Catalog 做為 Hive 中繼存放區，以在 中新增資料表中繼資料旁的方塊 AWS Glue Data Catalog。這會指定 AWS Glue Data Catalog 做為任務中所用 Data Catalog 資源的中繼存放區，並啟用 Lake Formation 許可，以便稍後套用至目錄資源。
6. 選擇 Save (儲存)。
7. 在動作下選擇執行。

   此任務會在lfdeltadb資料庫中建立名為 product的 Delta Lake 資料表。在 Lake Formation 主控台中驗證product資料表。

向 Lake Formation 註冊資料位置

接著，將 Amazon S3 路徑註冊為資料湖的根位置。

1. 開啟 Lake Formation 主控台，網址為資料湖管理員使用者 https://console.aws.amazon.com/lakeformation/：//www.。
2. 在導覽窗格的註冊和擷取下，選擇資料位置。
3. 在主控台的右上角，選擇註冊位置。
4. 在註冊位置頁面上，輸入下列內容：

   - Amazon S3 路徑 – 選擇瀏覽並選取 lf-otf-datalake-123456789012。按一下 Amazon S3 根位置旁的向右箭頭 (>) 以導覽至s3/buckets/lf-otf-datalake-123456789012/transactionaldata/native-delta位置。
   - IAM 角色 – 選擇 LF-OTF-RegisterRole作為 IAM 角色。
   - 選擇註冊位置。

在 Delta Lake 資料表上授予資料湖許可

在此步驟中，我們會將資料湖許可授予商業分析師使用者。

1. 在資料湖許可下，選擇授予。
2. 在授予資料許可畫面上，選擇 IAM 使用者和角色。

3. `lf-consumer-analystuser` 從下拉式清單。
4. 選擇具名資料目錄資源。
5. 針對資料庫，選擇 `lfdeltadb`。
6. 針對資料表，選擇 `product`。
7. 接下來，您可以指定資料欄來授予資料欄型存取。

    a. 在資料表許可下，選擇選取。

    b. 在資料許可下，選擇資料欄型存取，選擇包含資料欄。

    c. 選擇 `product_name`、 `price`和 `category`欄。

    d. 選擇 Grant (授予)。

使用 Athena 查詢 Delta Lake 資料表

現在開始查詢您使用 Athena 建立的 Delta Lake 資料表。如果您是第一次在 Athena 中執行查詢，則需要設定查詢結果位置。如需詳細資訊，請參閱指定查詢結果位置。

1. 以資料湖管理員使用者身分登出，並使用先前從 AWS CloudFormation 輸出記下的密碼，`BusinessAnalystUser`以美國東部（維吉尼亞北部）區域中的身分登入。
2. 前往 https://console.aws.amazon.com/athena/ 開啟 Athena 主控台。
3. 選擇設定，然後選取管理。
4. 在查詢結果的位置方塊中，輸入您在 AWS CloudFormation 輸出中建立的儲存貯體路徑。複製 `AthenaQueryResultLocation`(s3：//lf-otf-tutorial-123456789012/athena-results/) 和儲存的值。
5. 執行下列查詢以預覽存放在 Delta Lake 資料表中的 10 筆記錄：

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

    如需查詢 Delta Lake 資料表的詳細資訊，請參閱《Amazon Athena 使用者指南》中的查詢 Delta Lake 資料表一節。

步驟 5：清除 AWS 資源

清理資源

若要避免您的 產生不必要的費用 AWS 帳戶，請刪除您用於本教學課程 AWS 的資源。

1. 以 IAM 管理員的身分登入 AWS CloudFormation 主控台，網址為 http：//https://console.aws.amazon.com/cloudformation.healthnet.com。
2. 刪除雲端形成堆疊。您建立的資料表會自動與堆疊一起刪除。

使用 Lake Formation 標籤型存取控制管理資料湖

成千上萬的客戶正在建置 PB 級資料湖 AWS。其中許多客戶使用 AWS Lake Formation 在整個組織中輕鬆建置和共用其資料湖。隨著資料表和使用者的數目增加，資料管理員和管理員正在尋找方法來輕鬆大規模管理資料湖上的許可。Lake Formation 標籤型存取控制 (LF-TBAC) 可允許資料管理員建立 LF 標籤（根據其資料分類和本體），然後連接到資源，以解決此問題。

LF-TBAC 是一種授權策略，可根據屬性定義許可。在 Lake Formation 中，這些屬性稱為 LF 標籤。您可以將 LF 標籤連接至 Data Catalog 資源和 Lake Formation 主體。資料湖管理員可以使用 LF 標籤來指派和撤銷 Lake Formation 資源的許可。如需詳細資訊，請參閱 Lake Formation 標籤型存取控制。

本教學課程示範如何使用 AWS 公有資料集建立 Lake Formation 標籤型存取控制政策。此外，它示範如何查詢具有與其相關聯的 Lake Formation 標籤型存取政策的資料表、資料庫和資料欄。

您可以針對下列使用案例使用 LF-TBAC：

- 您有大量資料表和主體，資料湖管理員必須授予存取權
- 您想要根據 內部部署來分類資料，並根據分類來授予許可
- 資料湖管理員想要以鬆散耦合的方式動態指派許可

以下是使用 LF-TBAC 設定許可的高階步驟：

1. 資料管理員使用兩個 LF 標籤定義 標籤本體： `Confidential`和 `Sensitive`。使用 的資料`Confidential=True`具有更嚴格的存取控制。使用 的資料`Sensitive=True`需要分析師的特定分析。
2. 資料管理員會將不同的許可層級指派給資料工程師，以使用不同的 LF 標籤建置資料表。
3. 資料工程師會建置兩個資料庫： `tag_database`和 `col_tag_database`。中的所有資料表`tag_database`都使用 設定`Confidential=True`。中的所有資料表`col_tag_database`都已設定 `Confidential=False`。中資料表的某些資料欄`col_tag_database`會加上 標籤，`Sensitive=True`以滿足特定分析需求。
4. 資料工程師會授予分析師具有特定表達式條件 `Confidential=True`和 `Confidential=False`、的讀取許可`Sensitive=True`。

5. 使用此組態，資料分析師可以專注於使用正確的資料執行分析。

主題

- 目標對象
- 先決條件
- 步驟 1：佈建您的 資源
- 步驟 2：註冊您的資料位置、建立 LF 標籤本體，以及授予許可
- 步驟 3：建立 Lake Formation 資料庫
- 步驟 4：授予資料表許可
- 步驟 5：在 Amazon Athena 中執行查詢以驗證許可
- 步驟 6：清除 AWS 資源

目標對象

本教學課程適用於資料管理員、資料工程師和資料分析師。在 Lake Formation 中管理 AWS Glue Data Catalog 許可時，生產帳戶中的資料管理員會根據其支援的功能擁有功能擁有權，並且可以將存取權授予各種消費者、外部組織和帳戶。

下表列出本教學課程中使用的角色：

角色	描述
資料管理員（管理員）	`lf-data-steward` 使用者具有下列存取權： • 對 Data Catalog 中所有資源的讀取存取權 • 可以建立 LF 標籤並關聯至資料工程師角色，以便授予其他主體許可
資料工程師	`lf-data-engineer` 使用者具有下列存取權： • 完整讀取、寫入和更新對 Data Catalog 中所有資源的存取 • 資料湖中的資料位置許可

角色	描述
	• 可以建立 LF 標籤的關聯，並與 Data Catalog 建立關聯 • 可以將 LF 標籤連接至 資源，該資源會根據資料管理員建立的任何政策，提供主體存取權
資料分析	`lf-data-analyst` 使用者具有下列存取權： • 精細存取 Lake Formation 標籤型存取政策共用的資源

先決條件

開始本教學課程之前，您必須擁有可用於以具有正確許可的管理使用者身分登入 AWS 帳戶 的 。如需詳細資訊，請參閱完成初始 AWS 組態任務。

本教學課程假設您熟悉 IAM。如需 IAM 的相關資訊，請參閱 IAM 使用者指南。

步驟 1：佈建您的 資源

本教學課程包含快速設定的 AWS CloudFormation 範本。您可以檢閱和自訂它，以符合您的需求。範本會建立三個不同的角色（列於 中目標對象) 來執行此練習，並將 nyc-taxi-data 資料集複製到本機 Amazon S3 儲存貯體。

- Amazon S3 儲存貯體
- 適當的 Lake Formation 設定
- 適當的 Amazon EC2 資源
- 具有登入資料的三個 IAM 角色

建立您的 資源

1. 在美國東部（維吉尼亞北部）區域，登入位於 https://console.aws.amazon.com/cloudformation：//www./ 的 AWS CloudFormation 主控台。
2. 選擇啟動堆疊。
3. 選擇 Next (下一步)。

4. 在使用者組態區段中，輸入三個角色的密碼：`DataStewardUserPassword`、`DataEngineerUserPassword`和`DataAnalystUserPassword`。
5. 檢閱最終頁面上的詳細資訊，然後選取我確認 AWS CloudFormation 可能會建立 IAM 資源。
6. 選擇 Create (建立)。

   堆疊建立最多可能需要五分鐘。

> **Note**
>
> 完成教學課程後，您可能想要在 中刪除堆疊 AWS CloudFormation ，以避免繼續產生費用。確認資源在堆疊的事件狀態中已成功刪除。

步驟 2：註冊您的資料位置、建立 LF 標籤本體，以及授予許可

在此步驟中，資料管理員使用者使用兩個 LF 標籤定義標籤本體：`Confidential`和`Sensitive`，並讓特定 IAM 主體能夠將新建立的 LF 標籤連接至 資源。

註冊資料位置並定義 LF 標籤本體

1. 以資料管理員使用者 (`lf-data-steward`) 身分執行第一個步驟，以驗證 Amazon S3 中的資料和 Lake Formation 中的資料目錄。

   a. 登入位於 https://console.aws.amazon.com/lakeformation/ 的 Lake Formation 主控台，`lf-data-steward`如同部署 AWS CloudFormation 堆疊時使用的密碼。

   b. 在導覽窗格的 Permissionş 下，選擇管理角色和任務。

   c. 在資料湖管理員區段中選擇新增。

   d. 在新增管理員頁面上，針對 IAM 使用者和角色，選擇使用者 `lf-data-steward`。

   e. 選擇儲存以新增`lf-data-steward`為 Lake Formation 管理員。
2. 接下來，更新 Data Catalog 設定以使用 Lake Formation 許可來控制目錄資源，而不是以 IAM 為基礎的存取控制。

   a. 在導覽窗格的管理下，選擇 Data Catalog 設定。

   b. 取消勾選 僅對新資料庫使用 IAM 存取控制。

   c. 取消勾選 僅對新資料庫中的新資料表使用 IAM 存取控制。

d. 按一下 Save (儲存)。

3. 接著，註冊資料湖的資料位置。

a. 在導覽窗格的管理下，選擇資料湖位置。

b. 選擇註冊位置。

c. 在註冊位置頁面上，針對 Amazon S3 路徑輸入 `s3://lf-tagbased-demo-`*`Account-ID`*。

d. 對於 IAM 角色，保持預設值`AWSServiceRoleForLakeFormationDataAccess`不變。

e. 選擇 Lake Formation 作為許可模式。

f. 選擇註冊位置。

4. 接下來，定義 LF 標籤來建立本體。

a. 在導覽窗格中的許可下，選擇 LF 標籤和許可。

b. 選擇新增 LF 標籤。

c. 在 Key (索引鍵) 欄位，輸入 `Confidential`。

d. 針對值，新增 `True`和 `False`。

e. 選擇新增 LF 標籤。

f. 重複這些步驟以建立值`Sensitive`為 的 LF 標籤`True`。

您已建立此練習的所有必要 LF 標籤。

將許可授予 IAM 使用者

1. 接著，讓特定 IAM 主體能夠將新建立的 LF 標籤連接至資源。

a. 在導覽窗格中的許可下，選擇 LF 標籤和許可。

b. 在 LF-Tag 許可區段中，選擇授予許可。

c. 針對許可類型，選擇 LF-Tag 鍵/值對許可。

d. 選取 IAM 使用者和角色。

e. 對於 IAM 使用者和角色，搜尋並選擇`lf-data-engineer`角色。

f. 在 LF-Tags 區段中，新增`Confidential`值為 `True`的索引鍵`False`，以及值`key`為 的 `Sensitive` `True`。

g. 在許可下，選取描述和關聯許可和可授予許可。

h. 選擇 Grant (授予)。

2. 接下來，將許可授予 `lf-data-engineer`，以在 Data Catalog 和 建立的基礎 Amazon S3 儲存貯體上建立資料庫 AWS CloudFormation。

   a. 在導覽窗格中的管理下，選擇管理角色和任務。

   b. 在資料庫建立者區段中，選擇授予。

   c. 針對 IAM 使用者和角色，選擇`lf-data-engineer`角色。

   d. 針對目錄許可，選取建立資料庫。

   e. 選擇 Grant (授予)。

3. 接著，將 Amazon S3 儲存貯體的許可授予(`s3://lf-tagbased-demo-`*`Account-ID`*`)lf-data-engineer`使用者。

   a. 在導覽窗格的許可下，選擇資料位置。

   b. 選擇 Grant (授予)。

   c. 選取我的帳戶。

   d. 針對 IAM 使用者和角色，選擇`lf-data-engineer`角色。

   e. 針對儲存位置，輸入範本 建立的 Amazon S3 AWS CloudFormation 儲存貯體(`s3://lf-tagbased-demo-`*`Account-ID`*`)`。

   f. 選擇 Grant (授予)。

4. 接著，授予與 LF-Tag 表達式 相關聯之資源的可`lf-data-engineer`授予許可`Confidential=True`。

   a. 在導覽窗格的許可下，選擇資料湖許可。

   b. 選擇 Grant (授予)。

   c. 選取 IAM 使用者和角色。

   d. 選擇角色 `lf-data-engineer`。

   e. 在 LF 標籤或目錄資源區段中，選取符合 LF 標籤的資源。

   f. 選擇新增 LF 標籤鍵/值對。

   g. 新增`Confidential`值為 的金鑰`True`。

   h. 在資料庫許可區段中，選取描述資料庫許可和可授予許可。

   i. 在資料表許可區段中，針對資料表許可和可授予許可，選取描述、選取和修改。

   j. 選擇 Grant (授予)。

5. 接著，授予與 LF-Tag 表達式 相關聯之資源的可`lf-data-engineer`授予許可`Confidential=False`。

   a. 在導覽窗格的許可下，選擇資料湖許可。

   b. 選擇 Grant (授予)。

   c. 選取 IAM 使用者和角色。

   d. 選擇角色 `lf-data-engineer`。

   e. 選取符合 LF 標籤的資源。

   f. 選擇新增 LF 標籤。

   g. 新增值`Confidential`為 的金鑰`False`。

   h. 在資料庫許可區段中，選取描述資料庫許可和可授予許可。

   i. 在資料表和資料欄許可區段中，請勿選取任何項目。

   j. 選擇 Grant (授予)。

6. 接下來，我們會授予與 LF-Tag 鍵/值對`Confidential=False`和 相關聯的資源的可`lf-data-engineer`授予許可`Sensitive=True`。

   a. 在導覽窗格的許可下，選擇資料許可。

   b. 選擇 Grant (授予)。

   c. 選取 IAM 使用者和角色。

   d. 選擇角色 `lf-data-engineer`。

   e. 在 LF 標籤或目錄資源區段下，選取符合 LF 標籤的資源。

   f. 選擇新增 LF 標籤。

   g. 新增值`Confidential`為 的金鑰`False`。

   h. 選擇新增 LF 標籤鍵/值對。

   i. 新增值`Sensitive`為 的金鑰`True`。

   j. 在資料庫許可區段中，選取描述資料庫許可和可授予許可。

   k. 在資料表許可區段中，針對資料表許可和可授予許可，選取描述、選取和修改。

   l. 選擇 Grant (授予)。

步驟 3：建立 Lake Formation 資料庫

在此步驟中，您會建立兩個資料庫，並將 LF 標籤連接至資料庫和特定資料欄，以供測試之用。

建立資料庫和資料表以進行資料庫層級存取

1. 首先，建立資料庫 tag_database、資料表 source_data，並連接適當的 LF 標籤。

   a. 在 Lake Formation 主控台 (https://console.aws.amazon.com/lakeformation/) 的資料目錄下，選擇資料庫。

   b. 選擇建立資料庫。

   c. 對於名稱，輸入 tag_database。

   d. 針對位置，輸入範本 建立的 Amazon S3 AWS CloudFormation 位置(s3://lf-tagbased-demo-*Account-ID*/tag_database/)。

   e. 取消選取 僅針對此資料庫中的新資料表使用 IAM 存取控制。

   f. 選擇建立資料庫。

2. 接著，在 中建立新的資料表tag_database。

   a. 在資料庫頁面上，選取資料庫 tag_database。

   b. 選擇檢視資料表，然後按一下建立資料表。

   c. 對於名稱，輸入 source_data。

   d. 在 Database (資料庫) 中，選擇 tag_database 資料庫。

   e. 針對資料表格式，選擇標準 AWS Glue 資料表。

   f. 對於位於 的資料，選取我帳戶中的指定路徑。

   g. 針對包含路徑，輸入 AWS CloudFormation 範本 所tag_database建立的路徑(s3://lf-tagbased-demo*Account-ID*/tag_database/)。

   h. 針對資料格式，選取 CSV。

   i. 在上傳結構描述下，輸入下列資料欄結構的 JSON 陣列來建立結構描述：

```
[
              {
                   "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. 選擇上傳。上傳結構描述後，資料表結構描述看起來應該如下螢幕擷取畫面：

#	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. 選擇提交。

3. 接著，在資料庫層級連接 LF 標籤。

a. 在資料庫頁面上，尋找並選取 tag_database。

b. 在動作功能表中，選擇編輯 LF 標籤。

c. 選擇指派新的 LF 標籤。

d. 針對指派的 key，選擇您先前建立的 Confidential LF 標籤。

e. 針對值，選擇 True。

f. 選擇 Save (儲存)。

如此即完成 tag_database 資料庫的 LF-Tag 指派。

建立資料庫和資料表以進行資料欄層級存取

重複下列步驟來建立資料庫col_tag_database和資料表 source_data_col_lvl，並在資料欄層級連接 LF 標籤。

1. 在資料庫頁面上，選擇建立資料庫。
2. 對於名稱，輸入 col_tag_database。
3. 針對位置，輸入範本 建立的 Amazon S3 AWS CloudFormation 位置(s3://lf-tagbased-demo-*Account-ID*/col_tag_database/)。
4. 取消選取 僅針對此資料庫中的新資料表使用 IAM 存取控制。
5. 選擇建立資料庫。
6. 在資料庫頁面上，選取您的新資料庫 (col_tag_database)。
7. 選擇檢視資料表，然後按一下建立資料表。
8. 對於名稱，輸入 source_data_col_lvl。
9. 針對資料庫，選擇您的新資料庫 (col_tag_database)。
10. 針對資料表格式，選擇標準 AWS Glue 資料表。
11. 對於位於 的資料，選取我帳戶中的指定路徑。
12. 輸入 的 col_tag_database Amazon S3 路徑(s3://lf-tagbased-demo-*Account-ID*/col_tag_database/)。
13. 針對資料格式，選取 CSV。
14. 在 下Upload schema，輸入下列結構描述 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. 選擇 Upload。上傳結構描述後，資料表結構描述看起來應該如下螢幕擷取畫面。

#	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. 選擇提交以完成資料表的建立。
17. 現在，將 Sensitive=True LF 標籤與資料欄 vendorid和 建立關聯fare_amount。
    a. 在資料表頁面上，選取您建立的資料表(source_data_col_lvl)。
    b. 在動作功能表中，選擇結構描述。
    c. 選取資料欄vendorid，然後選擇編輯 LF 標籤。
    d. 針對指派的金鑰，選擇敏感。
    e. 針對值，選擇 True。
    f. 選擇 Save (儲存)。
18. 接著，將 Confidential=False LF 標籤與 建立關聯col_tag_database。這是必要項目，lf-data-analyst讓 能夠在登入col_tag_database時描述資料庫 Amazon Athena。
    a. 在資料庫頁面上，尋找並選取 col_tag_database。
    b. 在動作功能表中，選擇編輯 LF 標籤。
    c. 選擇指派新的 LF 標籤。
    d. 針對指派的金鑰，選擇您先前建立的 Confidential LF 標籤。
    e. 針對值，選擇 False。
    f. 選擇 Save (儲存)。

步驟 4：授予資料表許可

col_tag_database 使用 LF 標籤tag_database和 將許可授予資料分析師以取用資料庫Confidential和資料表Sensitive。

1. 請依照下列步驟，在與 LF-Tag Confidential=True(Database：tag_database) 相關聯的物件上將許可授予lf-data-analyst使用者，以取得資料表的Describe資料庫和Select許可。
    a. 以 登入 Lake Formation 主控台，網址為 https：//https://console.aws.amazon.com/lakeformation/.healthnet.comlf-data-engineer。
    b. 在許可下，選取資料湖許可。
    c. 選擇 Grant (授予)。
    d. 在主體下，選取 IAM 使用者和角色。
    e. 針對 IAM 使用者和角色，選擇 lf-data-analyst。
    f. 在 LF 標籤或目錄資源下，選取符合 LF 標籤的資源。

g. 選擇新增 LF 標籤。

h. 針對金鑰，選擇 Confidential。

i. 針對值，選擇 True。

j. 針對資料庫許可，選取 Describe。

k. 針對資料表許可，選擇選取和描述。

l. 選擇 Grant (授予)。

2. 接著，重複這些步驟，將 LF-Tag 表達式的許可授予資料分析師Confidential=False。從 Amazon Athena 以 身分登入source_data_col_lvl時，此 LF 標籤用於描述 lf-data-analyst col_tag_database和 資料表。

a. 以 登入 Lake Formation 主控台，網址為 http：/https://console.aws.amazon.com/lakeformation/.healthnet.comlf-data-engineer。

b. 在資料庫頁面上，選取資料庫 col_tag_database。

c. 選擇動作和授予。

d. 在主體下，選取 IAM 使用者和角色。

e. 針對 IAM 使用者和角色，選擇 lf-data-analyst。

f. 選取符合 LF 標籤的資源。

g. 選擇新增 LF 標籤。

h. 針對金鑰，選擇 Confidential。

i. 針對 Valueş，選擇 False。

j. 針對資料庫許可，選取 Describe。

k. 對於資料表許可，請勿選取任何項目。

l. 選擇 Grant (授予)。

3. 接下來，重複這些步驟，將 Confidential=False和 的 LF-Tag 表達式許可授予資料分析師Sensitive=True。從 Amazon Athena 以 身分登入時，此 LF 標籤用於描述 col_tag_database和 資料表 lf-data-analyst source_data_col_lvl（資料欄層級）。

a. 以 身分登入位於 https://console.aws.amazon.com/lakeformation/ 的 Lake Formation 主控台lf-data-engineer。

b. 在資料庫頁面上，選取資料庫 col_tag_database。

c. 選擇動作和授予。

d. 在主體下，選取 IAM 使用者和角色。

e. 針對 IAM 使用者和角色，選擇 lf-data-analyst。
f. 選取符合 LF 標籤的資源。
g. 選擇新增 LF 標籤。
h. 針對金鑰，選擇 Confidential。
i. 針對 Valueş，選擇 False。
j. 選擇新增 LF 標籤。
k. 針對金鑰，選擇 Sensitive。
l. 針對 Valueş，選擇 True。
m. 針對資料庫許可，選取 Describe。
n. 針對資料表許可，選取 Select和 Describe。
o. 選擇 Grant (授予)。

步驟 5：在 Amazon Athena 中執行查詢以驗證許可

在此步驟中，使用 Amazon Athena 對兩個資料表 執行SELECT查詢(source_data and source_data_col_lvl)。使用 Amazon S3 路徑做為查詢結果位置 (s3://lf-tagbased-demo-*Account-ID*/athena-results/)。

1. 以 身分登入位於 https://console.aws.amazon.com/athena/：//www.healthnet.com 的 Athena 主控台lf-data-analyst。
2. 在 Athena 查詢編輯器中，選擇左側面板tag_database中的 。
3. 選擇旁邊的其他功能表選項圖示 （ 三個垂直點 ）source_data，然後選擇預覽資料表。
4. 選擇 Run query (執行查詢)。

   查詢應該需要幾分鐘的時間來執行。查詢會顯示輸出中的所有資料欄，因為 LF 標籤在資料庫層級相關聯，且source_data資料表會自動LF-tag從資料庫繼承 tag_database。
5. 使用 col_tag_database和 執行另一個查詢source_data_col_lvl。

   第二個查詢會傳回標記為 Non-Confidential和 的兩個資料欄Sensitive。
6. 您也可以檢查，查看您沒有政策授予之資料欄上的 Lake Formation 標籤型存取政策行為。從資料表 選取未標記的資料欄時source_data_col_lvl，Athena 會傳回錯誤。例如，您可以執行下列查詢來選擇未標記的資料欄 geolocationid：

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

步驟 6：清除 AWS 資源

若要避免您的 產生不必要的費用 AWS 帳戶，您可以刪除您用於本教學課程 AWS 的資源。

1. 以 身分登入 Lake Formation 主控台，`lf-data-engineer`並刪除資料庫`tag_database`和`col_tag_database`。
2. 接下來，以 身分登入`lf-data-steward`並清除上述授予的所有 LF 標籤許可、資料許可和資料位置許可，這些許可已授予 `lf-data-engineer`和 `lf-data-analyst.`。
3. 使用您用來部署 AWS CloudFormation 堆疊的 IAM 登入資料，以帳戶擁有者的身分登入 Amazon S3 主控台。
4. 刪除下列儲存貯體：
   - lf-tagbased-demo-accesslogs-*acct-id*
   - lf-tagbased-demo-*acct-id*
5. 登入 AWS CloudFormation 主控台，網址為 https://console.aws.amazon.com/cloudformation：// www.healthnet.com/healthnet.com/com；；；，然後刪除您建立的堆疊。等待堆疊狀態變更為 `DELETE_COMPLETE`。

使用資料列層級存取控制保護資料湖

AWS Lake Formation 資料列層級許可可讓您根據資料合規和管理政策，提供資料表中特定資料列的存取權。如果您有存放數十億筆記錄的大型資料表，您需要一種方法來讓不同的使用者和團隊僅存取他們允許查看的資料。資料列層級存取控制是保護資料的簡單且高效能的方式，同時讓使用者存取執行其任務所需的資料。Lake Formation 透過識別哪些主體存取了哪些資料、何時以及透過哪些服務，提供集中式稽核和合規報告。

在本教學課程中，您將了解資料列層級存取控制如何在 Lake Formation 中運作，以及如何設定它們。

本教學課程包含 AWS CloudFormation 範本，可讓您快速設定所需的資源。您可以檢閱和自訂它，以符合您的需求。

主題

- 目標對象
- 先決條件
- 步驟 1：佈建您的 資源
- 步驟 2：不含資料篩選條件的查詢

- 步驟 3：設定資料篩選條件並授予許可
- 步驟 4：使用資料篩選條件查詢
- 步驟 5：清除 AWS 資源

目標對象

本教學課程適用於資料管理員、資料工程師和資料分析師。下表列出資料擁有者和資料取用者的角色和責任。

角色	描述
IAM 管理員	可建立使用者和角色以及 Amazon Simple Storage Service (Amazon S3) 儲存貯體的使用者。具有 `AdministratorAccess` AWS 受管政策。
資料湖管理員	負責設定資料湖、建立資料篩選條件，以及授予許可給資料分析師的使用者。
資料分析	可對資料湖執行查詢的使用者。居住在不同國家/地區的資料分析師 （針對我們的使用案例，美國和日本） 只能分析位於自己國家/地區的客戶的產品評論，而且基於合規原因， 應該無法查看位於其他國家/地區的客戶資料。

先決條件

開始本教學課程之前，您必須擁有可用於以具有正確許可的管理使用者身分登入 AWS 帳戶 的 。如需詳細資訊，請參閱完成初始 AWS 組態任務。

本教學課程假設您熟悉 IAM。如需 IAM 的相關資訊，請參閱 IAM 使用者指南。

變更 Lake Formation 設定

> **Important**
>
> 啟動 AWS CloudFormation 範本之前，請先依照下列步驟，停用在 Lake Formation 中僅對新資料庫/資料表使用 IAM 存取控制的選項：

1. 在美國東部（維吉尼亞北部）區域或美國西部（奧勒岡）區域，登入位於 https://console.aws.amazon.com/lakeformation/ 的 Lake Formation 主控台。
2. 在資料型錄下，選擇設定。
3. 取消選取 僅對新資料庫使用 IAM 存取控制，並僅對新資料庫中的新資料表使用 IAM 存取控制。
4. 選擇 Save (儲存)。

步驟 1：佈建您的 資源

本教學課程包含快速設定的 AWS CloudFormation 範本。您可以檢閱和自訂它，以符合您的需求。AWS CloudFormation 範本會產生下列資源：

- 使用者和政策：
  - DataLakeAdmin
  - DataAnalystUS
  - DataAnalystJP
- Lake Formation 資料湖設定和許可
- Lambda 函數（適用於 Lambda 支援的 AWS CloudFormation 自訂資源），用於將範例資料檔案從公有 Amazon S3 儲存貯體複製到 Amazon S3 儲存貯體
- 做為資料湖的 Amazon S3 儲存貯體
- AWS Glue Data Catalog 資料庫、資料表和分割區

建立您的 資源

請依照下列步驟，使用 AWS CloudFormation 範本建立您的 資源。

1. 在美國東部（維吉尼亞北部）區域，登入位於 https://console.aws.amazon.com/cloudformation：//www./ 的 AWS CloudFormation 主控台。

2. 選擇啟動堆疊。
3. 在建立堆疊畫面上選擇下一步。
4. 輸入堆疊名稱。
5. 針對 DatalakeAdminUserName 和 DatalakeAdminUserPassword，輸入資料湖管理員使用者的 IAM 使用者名稱和密碼。
6. 針對 DataAnalystUsUserName 和 DataAnalystUsUserPassword，輸入您要用於負責美國市場之資料分析師使用者的使用者名稱和密碼。
7. 針對 DataAnalystJpUserName 和 DataAnalystJpUserPassword，輸入您要用於負責日本市場之資料分析師使用者的使用者名稱和密碼。
8. 針對 DataLakeBucketName，輸入資料儲存貯體的名稱。
9. 對於 DatabaseName 和 TableName，保留 為預設值。
10. 選擇下一步
11. 在下一頁中，選擇下一步。
12. 檢閱最終頁面上的詳細資訊，然後選取我確認 AWS CloudFormation 可能會建立 IAM 資源。
13. 選擇 Create (建立)。

    堆疊建立可能需要一分鐘才能完成。

步驟 2：不含資料篩選條件的查詢

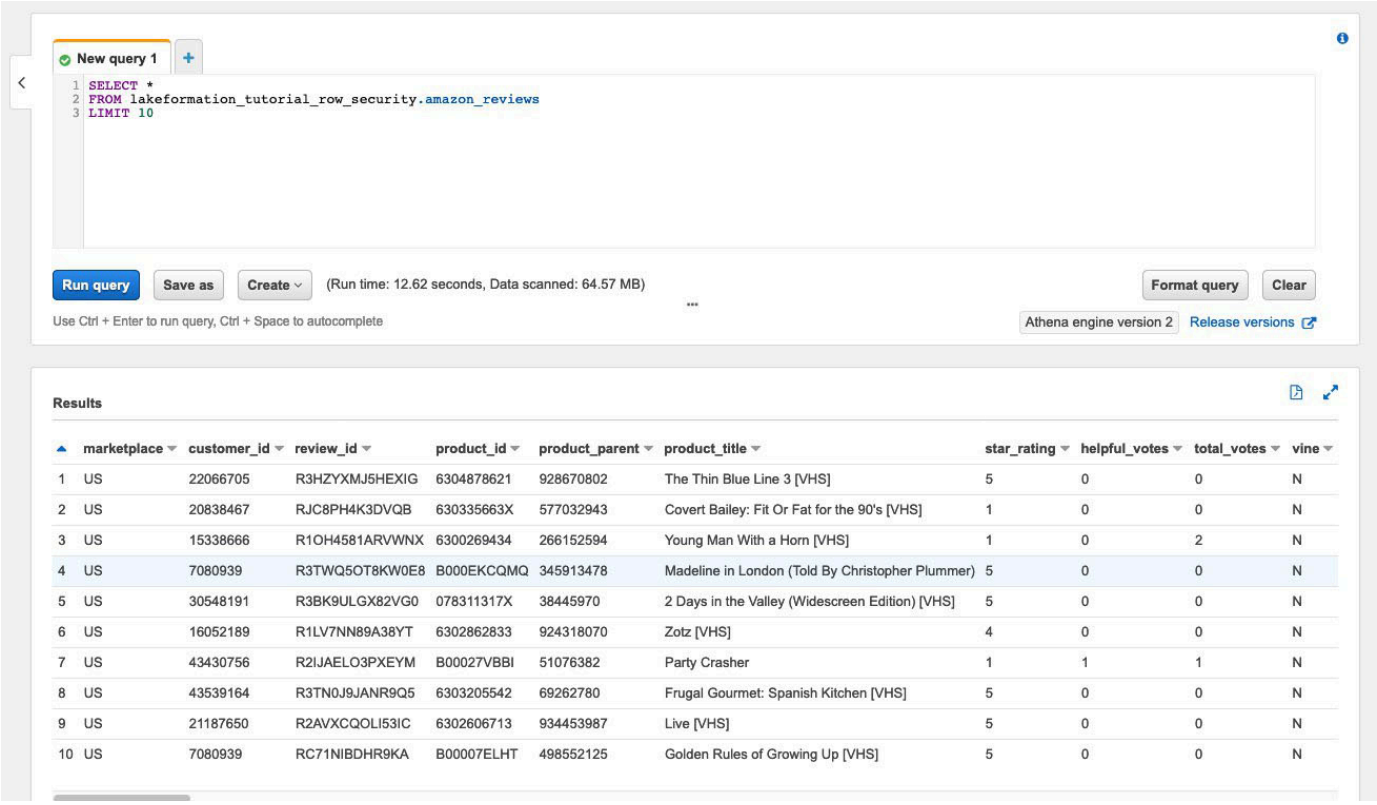
設定環境後，您可以查詢產品評論資料表。首先查詢沒有資料列層級存取控制的資料表，以確保您可以看到資料。如果您是第一次在 Amazon Athena 中執行查詢，則需要設定查詢結果位置。

在沒有資料列層級存取控制的情況下查詢資料表

1. 以 `DatalakeAdmin` 使用者https://console.aws.amazon.com/athena/身分登入 Athena主控台，並執行下列查詢：

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

   下列螢幕擷取畫面顯示查詢結果。此資料表只有一個分割區 `product_category=Video`，因此每個記錄都是影片產品的檢閱註解。

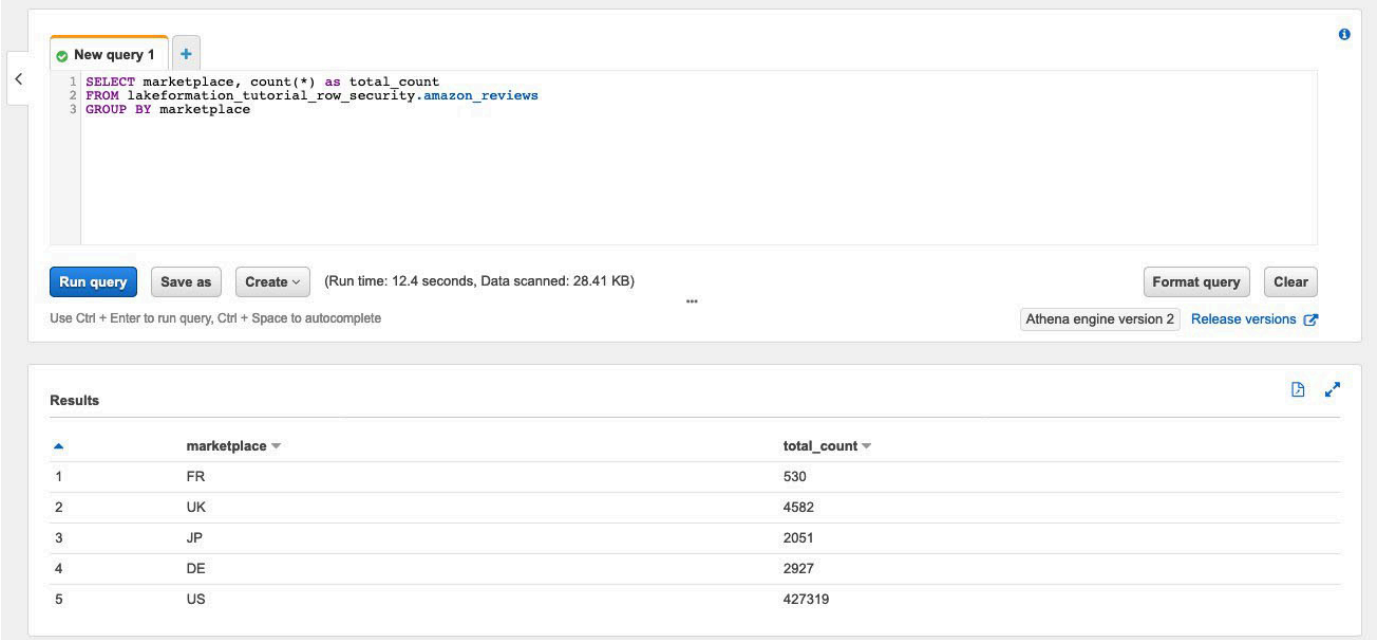


	marketplace	customer_id	review_id	product_id	product_parent	product_title	star_rating	helpful_votes	total_votes	vine
1	US	22066705	R3HZYXMJ5HEXIG	6304878621	928670802	The Thin Blue Line 3 [VHS]	5	0	0	N
2	US	20838467	RJC8PH4K3DVQB	630335663X	577032943	Covert Bailey: Fit Or Fat for the 90's [VHS]	1	0	0	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. 接著，執行彙總查詢，擷取每個 的記錄總數marketplace。

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

下列螢幕擷取畫面顯示查詢結果。資料marketplace欄有五個不同的值。在後續步驟中，您將使用 marketplace 欄設定資料列型篩選條件。



步驟 3：設定資料篩選條件並授予許可

本教學課程使用兩個資料分析師：一個負責美國市場，另一個負責日本市場。每個分析師都使用 Athena 來分析客戶評論，僅限其特定市場。建立兩個不同的資料篩選條件，一個用於負責美國市場的分析師，另一個則用於負責日本市場的分析師。然後，授予分析師各自的許可。

建立資料篩選條件並授予許可

1. 建立篩選條件以限制對USmarketplace資料的存取。

   a. 以 DatalakeAdmin使用者身分登入美國東部（維吉尼亞北部）區域的 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

   b. 選擇資料篩選條件。

   c. 選擇建立新篩選條件。

   d. 針對資料篩選條件名稱，輸入 amazon_reviews_US。

   e. 針對目標資料庫，選擇資料庫 lakeformation_tutorial_row_security。

   f. 針對目標資料表，選擇資料表 amazon_reviews。

   g. 對於資料欄層級存取，請將 保留為預設值。

   h. 針對資料列篩選條件表達式，輸入 marketplace='US'。

   i. 選擇 Create filter (建立篩選條件)。

2. 建立篩選條件以限制對日文marketplace資料的存取。

a. 在資料篩選條件頁面上，選擇建立新篩選條件。

b. 針對資料篩選條件名稱，輸入 amazon_reviews_JP。

c. 針對目標資料庫，選擇資料庫 lakeformation_tutorial_row_security。

d. 針對目標資料表，選擇 table amazon_reviews。

e. 對於資料欄層級存取，請將 保留為預設值。

f. 針對資料列篩選條件表達式，輸入 marketplace='JP'。

g. 選擇 Create filter (建立篩選條件)。

3. 接著，使用這些資料篩選條件將許可授予資料分析師。請依照下列步驟將許可授予美國資料分析師 (DataAnalystUS)：

a. 在許可下，選擇資料湖許可。

b. 在資料許可下，選擇授予。

c. 針對主體，選擇 IAM 使用者和角色，然後選取角色 DataAnalystUS。

d. 針對 LF 標籤或目錄資源，選擇具名資料目錄資源。

e. 針對 Database (資料庫)，輸入 lakeformation_tutorial_row_security。

f. 對於資料表 - 選用，選擇 amazon_reviews。

g. 針對資料篩選條件 – 選用 選取 amazon_reviews_US。

h. 針對資料篩選條件許可，選取選取。

i. 選擇 Grant (授予)。

4. 請依照下列步驟將許可授予日文資料分析師 (DataAnalystJP)：

a. 在許可下，選擇資料湖許可。

b. 在資料許可下，選擇授予。

c. 針對主體，選擇 IAM 使用者和角色，然後選取角色 DataAnalystJP。

d. 針對 LF 標籤或目錄資源，選擇具名資料目錄資源。

e. 針對 Database (資料庫)，輸入 lakeformation_tutorial_row_security。

f. 對於資料表 - 選用，選擇 amazon_reviews。

g. 針對資料篩選條件 – 選用 選取 amazon_reviews_JP。

h. 針對資料篩選條件許可，選取選取。

i. 選擇 Grant (授予)。

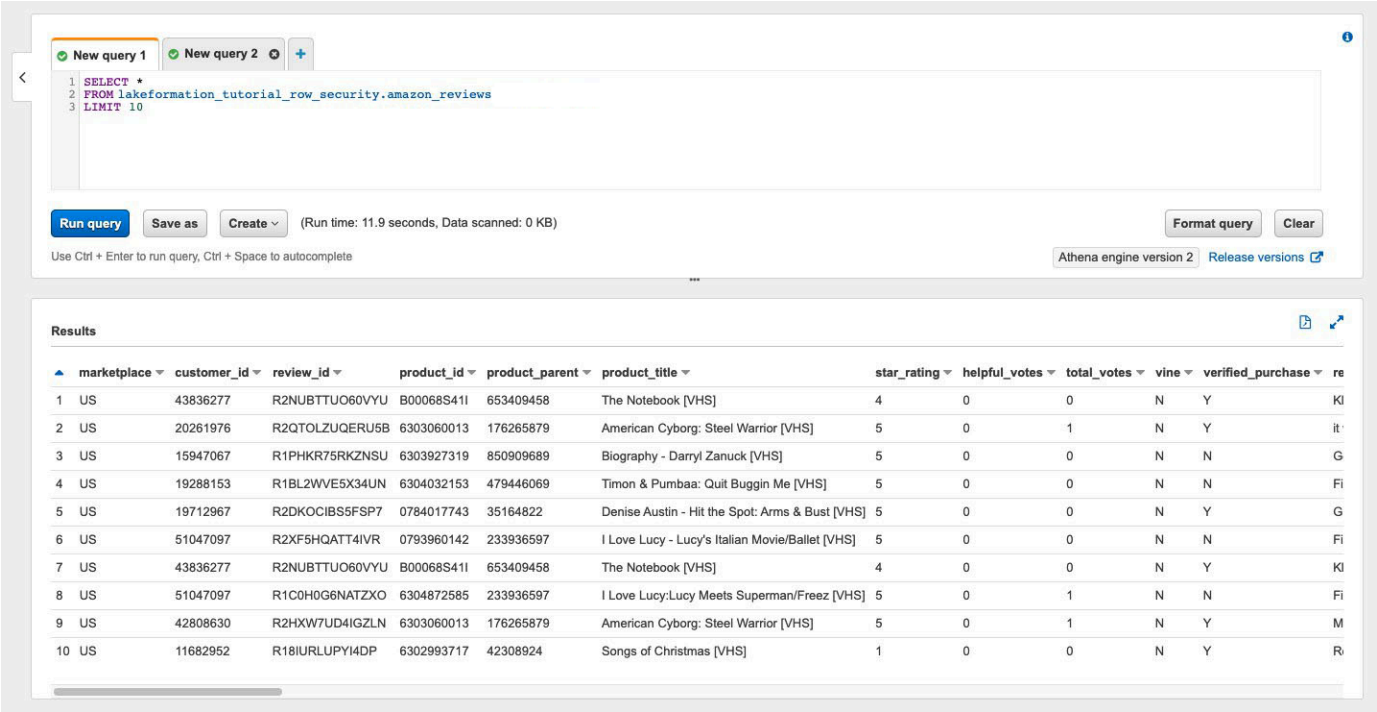
步驟 4：使用資料篩選條件查詢

將資料篩選條件連接至產品檢閱資料表後，請執行一些查詢，並查看 Lake Formation 如何強制執行許可。

1. 以 DataAnalystUS 使用者身分登入位於 https://console.aws.amazon.com/athena/：// 的 Athena 主控台。
2. 執行下列查詢以擷取一些記錄，這些記錄會根據我們定義的資料列層級許可進行篩選：

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

下列螢幕擷取畫面顯示查詢結果。



	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. 同樣地，請執行查詢來計算每個市集的記錄總數。

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

查詢結果只會在結果marketplaceUS中顯示 。這是因為使用者只能查看marketplace資料欄值等於 的資料列US。

4. 切換到DataAnalystJP使用者並執行相同的查詢。

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

查詢結果只會顯示屬於 JP 的記錄marketplace。

5. 執行查詢以計算每個 的記錄總數marketplace。

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

查詢結果只會顯示屬於 JP 的資料列marketplace。

步驟 5：清除 AWS 資源

清除資源

若要避免您的 產生不必要的費用 AWS 帳戶，您可以刪除您用於本教學課程 AWS 的資源。

- 刪除雲端形成堆疊。

使用 Lake Formation 標籤型存取控制和具名資源共用資料湖

本教學課程示範如何設定 AWS Lake Formation ，以安全地與多家公司、組織或業務單位共用存放在資料湖中的資料，而不必複製整個資料庫。使用 Lake Formation 跨帳戶存取控制 AWS 帳戶 ，有兩個選項可以與另一個資料庫和資料表共用：

- Lake Formation 標籤型存取控制 （ 建議 ）

  Lake Formation 標籤型存取控制是一種授權策略，可根據屬性定義許可。在 Lake Formation 中，這些屬性稱為 LF 標籤。如需更多詳細資訊，請參閱 使用 Lake Formation 標籤型存取控制管理資料湖。
- Lake Formation 命名資源

Lake Formation 命名的資源方法是一種授權策略，可定義資源的許可。資源包括資料庫、資料表和資料欄。資料湖管理員可以指派和撤銷 Lake Formation 資源的許可。如需更多詳細資訊，請參閱 Lake Formation 中的跨帳戶資料共用。

如果資料湖管理員偏好將許可明確授予個別資源，建議使用具名資源。當您使用具名資源方法將 Data Catalog 資源的 Lake Formation 許可授予外部帳戶時，Lake Formation 會使用 AWS Resource Access Manager (AWS RAM) 來共用資源。

主題

- 目標對象
- 在生產者帳戶中設定 Lake Formation Data Catalog 設定
- 步驟 1：使用 AWS CloudFormation 範本佈建資源
- 步驟 2：Lake Formation 跨帳戶共用先決條件
- 步驟 3：使用標籤型存取控制方法實作跨帳戶共用
- 步驟 4：實作具名資源方法
- 步驟 5：清除 AWS 資源

目標對象

本教學課程適用於資料管理員、資料工程師和資料分析師。在 Lake Formation 中共用 Data Catalog 資料表 AWS Glue 和管理許可時，生產帳戶中的資料管理員會根據其支援的功能擁有功能所有權，並可以授予各種取用者、外部組織和帳戶的存取權。下表列出本教學課程中使用的角色：

角色	描述
DataLakeAdminProducer	資料湖管理員 IAM 使用者具有下列存取權： • 完整讀取、寫入和更新對 Data Catalog 中所有資源的存取 • 能夠授予 資源許可 • 可以建立共用資料表的資源連結 • 可以將 LF 標籤連接至 資源，該資源會根據資料管理員建立的任何政策，提供主體存取權

角色	描述
DataLakeAdminConsumer	資料湖管理員 IAM 使用者具有下列存取權： • 完整讀取、寫入和更新對 Data Catalog 中所有資源的存取 • 能夠授予 資源許可 • 可以建立共用資料表的資源連結 • 可以將 LF 標籤連接至 資源，該資源會根據資料管理員建立的任何政策，提供主體存取權
DataAnalyst	DataAnalyst 使用者具有下列存取權： • 精細存取 Lake Formation 標籤型存取政策或使用具名資源方法共用的資源

在生產者帳戶中設定 Lake Formation Data Catalog 設定

開始本教學課程之前，您必須擁有可用於以具有正確許可的管理使用者身分登入 AWS 帳戶 的 。如需詳細資訊，請參閱完成初始 AWS 組態任務。

本教學課程假設您熟悉 IAM。如需 IAM 的詳細資訊，請參閱 IAM 使用者指南。

在生產者帳戶中設定 Lake Formation Data Catalog 設定

> **Note**
>
> 在本教學課程中，具有來源資料表的帳戶稱為生產者帳戶，而需要存取來源資料表的帳戶稱為取用者帳戶。

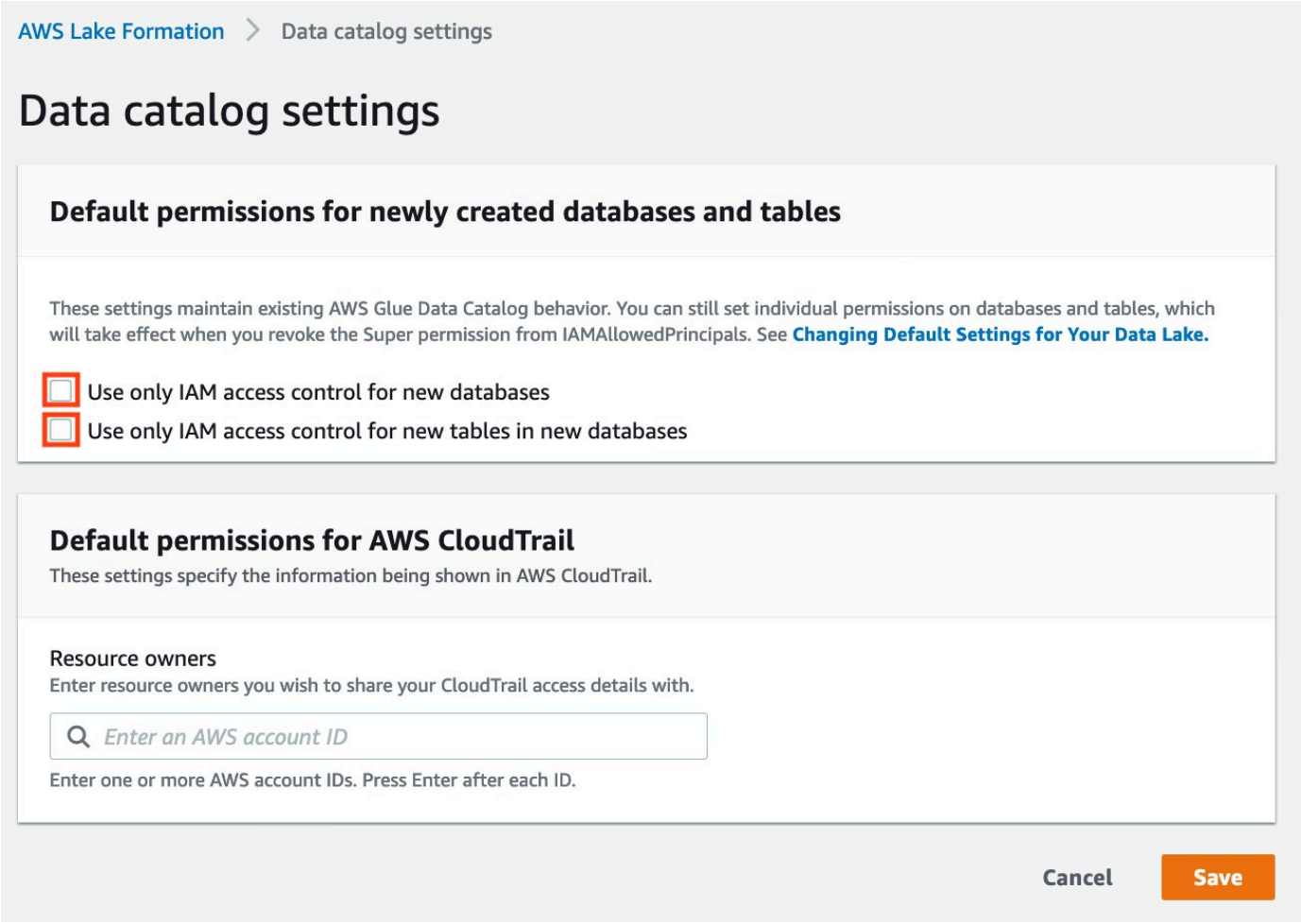
Lake Formation 提供自己的許可管理模型。為了維持與 IAM 許可模型的回溯相容性，預設會將 `Super`許可授予所有現有 AWS Glue Data Catalog 資源`IAMAllowedPrincipals`上的 群組。此外，對新的 Data Catalog 資源僅啟用 IAM 存取控制設定。本教學課程使用 Lake Formation 許可的精細存取控制，並使用 IAM 政策進行粗略精細存取控制。如需詳細資訊，請參閱 精細存取控制的方法。因此，在使用 AWS CloudFormation 範本進行快速設定之前，您需要變更生產者帳戶中的 Lake Formation Data Catalog 設定。

> **Important**
>
> 此設定會影響所有新建立的資料庫和資料表，因此我們強烈建議在非生產帳戶或新帳戶中完成此教學課程。此外，如果您使用的是共用帳戶（例如公司的開發帳戶），請確定它不會影響其他資源。如果您偏好保留預設安全設定，則必須在將資源分享給其他帳戶時完成額外的步驟，在該帳戶中，您必須撤銷資料庫或資料表`IAMAllowedPrincipals`上的預設超級許可。我們稍後會討論本教學課程的詳細資訊。

若要在生產者帳戶中設定 Lake Formation Data Catalog 設定，請完成下列步驟：

1. AWS Management Console 使用生產者帳戶以管理員使用者身分，或以具有 Lake Formation `PutDataLakeSettings` API 許可的使用者身分登入 。
2. 在 Lake Formation 主控台的導覽窗格中，於資料目錄下，選擇設定。
3. 取消選取 僅對新資料庫使用 IAM 存取控制，並僅對新資料庫中的新資料表使用 IAM 存取控制

   選擇 Save (儲存)。



此外，您可以在管理角色和任務、資料庫建立者IAMAllowedPrincipals下移除的CREATE_DATABASE許可。只有在那時，您才可以管理誰可以透過 Lake Formation 許可建立新的資料庫。

步驟 1：使用 AWS CloudFormation 範本佈建資源

生產者帳戶的 CloudFormation 範本會產生下列資源：

- 做為資料湖的 Amazon S3 儲存貯體。
- Lambda 函數（適用於 Lambda 支援的 AWS CloudFormation 自訂資源）。我們使用 函數將範例資料檔案從公有 Amazon S3 儲存貯體複製到您的 Amazon S3 儲存貯體。
- IAM 使用者和政策：DataLakeAdminProducer。
- 適當的 Lake Formation 設定和許可，包括：
  - 在生產者帳戶中定義 Lake Formation 資料湖管理員

  - 將 Amazon S3 儲存貯體註冊為 Lake Formation 資料湖位置 （生產者帳戶）
- AWS Glue Data Catalog 資料庫、資料表和分割區。由於有兩個共用資源的選項 AWS 帳戶，因此此範本會建立兩組不同的資料庫和資料表。

消費者帳戶的 AWS CloudFormation 範本會產生下列資源：

- IAM 使用者和政策：
  - DataLakeAdminConsumer
  - DataAnalyst
- AWS Glue Data Catalog 資料庫。此資料庫用於建立共用資源的資源連結。

在生產者帳戶中建立您的資源

1. 在美國東部（維吉尼亞北部）區域，登入位於 https://console.aws.amazon.com/cloudformation：//www./ 的 AWS CloudFormation 主控台。
2. 選擇啟動堆疊。
3. 選擇 Next (下一步)。
4. 針對堆疊名稱，輸入堆疊名稱，例如 `stack-producer`。
5. 在使用者組態區段中，輸入 和 `ProducerDatalakeAdminUserName` 的使用者名稱和密碼`ProducerDatalakeAdminUserPassword`。
6. 針對 DataLakeBucketName，輸入資料湖儲存貯體的名稱。此名稱必須是全域唯一的。
7. 針對 DatabaseName 和 TableName，保留預設值。
8. 選擇 Next (下一步)。
9. 在下一頁中，選擇下一步。
10. 檢閱最終頁面上的詳細資訊，然後選取我確認 AWS CloudFormation 可能會建立 IAM 資源。
11. 選擇 Create (建立)。

    堆疊建立最多可能需要一分鐘。

在消費者帳戶中建立您的資源

1. 在美國東部（維吉尼亞北部）區域，登入位於 https://console.aws.amazon.com/cloudformation：//www./ 的 AWS CloudFormation 主控台。
2. 選擇啟動堆疊。

3. 選擇 Next (下一步)。
4. 針對堆疊名稱，輸入堆疊名稱，例如 `stack-consumer`。
5. 在使用者組態區段中，輸入 和 ConsumerDatalakeAdminUserName 的使用者名稱和密碼ConsumerDatalakeAdminUserPassword。
6. 針對 DataAnalystUserName和 DataAnalystUserPassword，輸入您想要的資料分析師 IAM 使用者的使用者名稱和密碼。
7. 針對 DataLakeBucketName，輸入資料湖儲存貯體的名稱。此名稱必須是全域唯一的。
8. 針對 DatabaseName，保留預設值。
9. 針對 AthenaQueryResultS3BucketName，輸入存放 Amazon Athena 查詢結果的 Amazon S3 儲存貯體名稱。 Amazon Athena 如果您沒有儲存貯體，請建立 Amazon S3 儲存貯體。
10. 選擇 Next (下一步)。
11. 在下一頁中，選擇下一步。
12. 檢閱最終頁面上的詳細資訊，然後選取我確認 AWS CloudFormation 可能會建立 IAM 資源。
13. 選擇 Create (建立)。

    堆疊建立最多可能需要一分鐘。

> **Note**
>
> 完成教學課程後，請在 中刪除堆疊 AWS CloudFormation ，以避免產生費用。確認資源在堆疊的事件狀態中已成功刪除。

步驟 2：Lake Formation 跨帳戶共用先決條件

在與 Lake Formation 共用資源之前，標籤型存取控制方法和具名資源方法都有先決條件。

完整的標籤型存取控制跨帳戶資料共用先決條件

- 如需跨帳戶資料共用需求的詳細資訊，請參閱跨帳戶資料共用章節中的先決條件一節。

  若要與跨帳戶版本設定第 3 版或更新版本共用 Data Catalog 資源，授予者需要擁有AWSLakeFormationCrossAccountManager您帳戶中 AWS 受管政策中定義的 IAM 許可。

  如果您使用的是跨帳戶版本設定的第 1 版或第 2 版，則必須先將下列JSON許可物件新增至生產者帳戶中的 Data Catalog 資源政策，才能使用標籤型存取控制方法授予跨帳戶存取資源。這可讓消

費者帳戶在 glue:EvaluatedByLakeFormationTags 為 true 時存取 Data Catalog。此外，對於您使用 Lake Formation 許可標籤將許可授予消費者帳戶的資源，此條件會變成 true。 AWS 帳戶 您授予許可的每個 都需要此政策。

下列政策必須在 Statement元素內。我們在下一節討論完整的 IAM 政策。

```
{
    "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
        }
    }
}
```

完成具名資源方法跨帳戶共用先決條件

1. 如果您的帳戶中沒有 Data Catalog 資源政策，則 Lake Formation 跨帳戶授予您照常進行。不過，如果 Data Catalog 資源政策存在，您必須將下列陳述式新增至其中，以允許使用具名資源方法進行跨帳戶授予時成功。如果您打算僅使用具名資源方法，或僅使用標籤型存取控制方法，您可以略過此步驟。在本教學課程中，我們會評估這兩種方法，而且我們需要新增下列政策。

   下列政策必須在 Statement元素內。我們在下一節討論完整的 IAM 政策。

```
{
          "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. 接著，使用 AWS Command Line Interface () 新增 AWS Glue Data Catalog 資源政策AWS CLI。

   如果您同時使用標籤型存取控制方法和具名資源方法授予跨帳戶許可，則必須在新增上述政策時將EnableHybrid引數設定為「true」。由於 主控台目前不支援此選項，因此您必須使用 glue:PutResourcePolicy API 和 AWS CLI。

   首先，建立政策文件 （ 例如 policy.json) 並新增上述兩個政策。將 *consumer-account-id* 取代為 AWS 帳戶 接收授予*### ID*、將 *region* 取代為 Data Catalog 的區域，其中包含您授予許可的資料庫和資料表，以及將 *account-id* 取代為生產者 AWS 帳戶 ID。

```
{
    "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"
                }
            }
        }
    ]
}
```

輸入下列 AWS CLI 命令。將 *glue-resource-policy* 取代為正確的值 （ 例如 file：// policy.json)。

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

如需詳細資訊，請參閱 put-resource-policy。

步驟 3：使用標籤型存取控制方法實作跨帳戶共用

在本節中，我們會逐步引導您完成下列高階步驟：

1. 定義 LF 標籤。
2. 將 LF 標籤指派給目標資源。
3. 將 LF-Tag 許可授予取用者帳戶。
4. 將資料許可授予取用者帳戶。
5. 或者，撤銷資料庫、資料表和資料欄IAMAllowedPrincipals上的 許可。
6. 建立共用資料表的資源連結。
7. 建立 LF 標籤並將其指派給目標資料庫。
8. 將 LF-Tag 資料許可授予取用者帳戶。

定義 LF 標籤

> **Note**
>
> 如果您已登入生產者帳戶，請先登出，再完成下列步驟。

1. 以 資料湖管理員身分登入生產者帳戶，網址為 https：//https://console.aws.amazon.com/lakeformation/。使用您在堆疊建立期間 AWS CloudFormation 指定的生產者帳號、IAM 使用者名稱 （ 預設值為 `DatalakeAdminProducer`) 和密碼。
2. 在 Lake Formation 主控台 (https://console.aws.amazon.com/lakeformation/) 的導覽窗格中，於許可下，選擇 LF 標籤和許可。
3. 選擇新增 LF 標籤。

將 LF 標籤指派給目標資源

將 LF-Tag 指派給目標資源，並將資料許可授予另一個帳戶

身為資料湖管理員，您可以將標籤連接至資源。如果您打算使用單獨的角色，則可能需要授予描述許可，並將許可連接到單獨的角色。

1. 在導覽窗格中的資料目錄下，選取資料庫。
2. 選取目標資料庫(`lakeformation_tutorial_cross_account_database_tbac`)，然後在動作功能表上，選擇編輯 LF 標籤。

   在本教學課程中，您可以將 LF 標籤指派給資料庫，但您也可以將 LF 標籤指派給資料表和資料欄。
3. 選擇指派新的 LF 標籤。
4. 新增索引鍵 `Confidentiality` 和值 `public`。
5. 選擇 Save (儲存)。

將 LF-Tag 許可授予取用者帳戶

仍在生產者帳戶中，授予許可給取用者帳戶存取 LF 標籤。

1. 在導覽窗格的許可下，選擇 LF 標籤和許可。

2. 選擇 LF-Tags 索引標籤，然後選擇與消費者帳戶共用的 LF-Tag 金鑰和值 (金鑰Confidentiality和值 public)。
3. 選擇 授予許可。
4. 針對許可類型，選擇 LF-Tag 鍵/值對許可。
5. 針對委託人，選擇外部帳戶。
6. 輸入目標 AWS 帳戶 ID。

   AWS 帳戶 會自動出現在同一個組織內。否則，您必須手動輸入 AWS 帳戶 ID。
7. 在許可下，選取描述。

   這是提供給消費者帳戶的許可。可授予許可是消費者帳戶可授予其他委託人的許可。
8. 選擇 Grant (授予)。

   此時，消費者資料湖管理員應該能夠在許可、LF 標籤和許可下，找到透過消費者帳戶 Lake Formation 主控台共用的政策標籤。

將資料許可授予取用者帳戶

我們現在將透過指定 LF-Tag 表達式，並授予取用者帳戶存取符合表達式的任何資料表或資料庫，來提供資料存取權給取用者帳戶。

1. 在導覽窗格的許可、資料湖許可下，選擇授予。
2. 針對委託人，選擇外部帳戶，然後輸入目標 AWS 帳戶 ID。
3. 對於 LF 標籤或目錄資源，選擇與消費者帳戶共用的 LF 標籤的索引鍵和值 (索引鍵Confidentiality和值 public)。
4. 針對許可，在 LF-Tags 相符的資源 （ 建議 ） 下，選擇新增 LF-Tag。
5. 選取與消費者帳戶共用的標籤的索引鍵和值 （ 索引鍵Confidentiality和值 public)。
6. 針對資料庫許可，選取資料庫許可下的描述，以授予資料庫層級的存取許可。
7. 取用者資料湖管理員應該能夠在 Lake Formation 主控台的許可、管理角色和任務https://console.aws.amazon.com/lakeformation/、LF 標籤下，找到透過取用者帳戶共用的政策標籤。
8. 選取在可授予許可下描述，讓取用者帳戶可以將資料庫層級許可授予其使用者。
9. 針對資料表和資料欄許可，選取資料表許可下的選取和描述。
10. 選取可授予許可下的選取和描述。
11. 選擇 Grant (授予)。

在資料庫、資料表和資料欄**IAMAllowedPrincipals**上撤銷 的許可 （選用）。

在本教學課程一開始，您已變更 Lake Formation Data Catalog 設定。如果您略過該部分，則需要此步驟。如果您變更 Lake Formation Data Catalog 設定，您可以略過此步驟。

在此步驟中，我們需要撤銷資料庫或資料表IAMAllowedPrincipals上來自 的預設超級許可。如需詳細資訊，請參閱 步驟 4：將資料存放區切換至 Lake Formation 許可模型。

在撤銷 的許可之前IAMAllowedPrincipals，請確定您已透過 Lake Formation 授予現有 IAM 主體必要的許可。這包括三個步驟：

1. 使用 Lake Formation GetDataAccess動作 （使用 IAM 政策），將 IAM 許可新增至目標 IAM 使用者或角色。
2. 授予具有 Lake Formation 資料許可 （變更、選取等） 的目標 IAM 使用者或角色。
3. 然後，撤銷 的許可IAMAllowedPrincipals。否則，在撤銷 的許可之後IAMAllowedPrincipals，現有的 IAM 主體可能無法再存取目標資料庫或 Data Catalog。

   當您想要套用 Lake Formation 許可模型 （而非 IAM 政策模型），以使用 Lake Formation 許可模型管理單一帳戶內或多個帳戶之間的使用者存取權時，IAMAllowedPrincipals需要撤銷 的超級許可。您不需要撤銷其他資料表IAMAllowedPrincipals的 許可，而您要保留傳統的 IAM 政策模型。

   此時，消費者帳戶資料湖管理員應該能夠在 https://console.aws.amazon.com/lakeformation/：// 的 Lake Formation 主控台、Data Catalog 資料庫下，找到透過消費者帳戶共用的資料庫和資料表。如果沒有，請確認下列項目是否已正確設定：

   1. 正確的政策標籤和值會指派給目標資料庫和資料表。
   2. 將正確的標籤許可和資料許可指派給取用者帳戶。
   3. 在資料庫或資料表IAMAllowedPrincipals上從 撤銷預設超級許可。

建立共用資料表的資源連結

在帳戶之間共用資源時，共用資源不會放入消費者帳戶的資料目錄中。為了提供它們，並使用 Athena 等服務查詢共用資料表的基礎資料，我們需要建立共用資料表的資源連結。資源連結是 Data Catalog 物件，是本機或共用資料庫或資料表的連結。如需詳細資訊，請參閱 建立資源連結。透過建立資源連結，您可以：

- 將不同的名稱指派給與 Data Catalog 資源命名政策一致的資料庫或資料表。
- 使用 Athena 和 Redshift Spectrum 等服務來查詢共用資料庫或資料表。

若要建立資源連結，請完成下列步驟：

1. 如果您已登入您的消費者帳戶，請登出。
2. 以消費者帳戶資料湖管理員身分登入。使用您在 AWS CloudFormation 堆疊建立期間指定的消費者帳戶 ID、IAM 使用者名稱 （ 預設 DatalakeAdminConsumer) 和密碼。
3. 在 Lake Formation 主控台 (https://console.aws.amazon.com/lakeformation/) 的導覽窗格中，於 Data Catalog、資料庫下，選擇共用資料庫 `lakeformation_tutorial_cross_account_database_tbac`。

   如果您沒有看到資料庫，請重新檢視先前的步驟，以查看是否已正確設定所有項目。
4. 選擇檢視資料表。
5. 選擇共用資料表 `amazon_reviews_table_tbac`。
6. 在動作功能表中，選擇建立資源連結。
7. 針對資源連結名稱，輸入名稱 （ 本教學課程為 `amazon_reviews_table_tbac_resource_link`)。
8. 在資料庫下，選取資源連結建立所在的資料庫 （ 在此文章中，n AWS CloudFormation堆疊已建立資料庫 `lakeformation_tutorial_cross_account_database_consumer`)。
9. 選擇 Create (建立)。

   資源連結會顯示在資料目錄 資料表 下。

建立 LF 標籤並將其指派給目標資料庫

Lake Formation 標籤位於與 資源相同的 Data Catalog 中。這表示在生產者帳戶中建立的標籤在授予取用者帳戶中資源連結的存取權時無法使用。您需要在取用者帳戶中建立一組單獨的 LF 標籤，以便在共用取用者帳戶中的資源連結時使用 LF 標籤型存取控制。

1. 在消費者帳戶中定義 LF 標籤。在本教學課程中，我們使用金鑰`Division`和值 `sales`、`marketing`和 `analyst`。
2. 將 LF 標籤索引鍵`Division`和值`analyst`指派給建立資源連結`lakeformation_tutorial_cross_account_database_consumer`的資料庫 。

將 LF 標籤資料許可授予消費者

最後一步，將 LF 標籤資料許可授予消費者。

1. 在導覽窗格的許可、資料湖許可下，選擇授予。
2. 針對主體，選擇 IAM 使用者和角色，然後選擇使用者 DataAnalyst。
3. 針對 LF 標籤或目錄資源，選擇符合 LF 標籤的資源 （建議）。
4. 選擇金鑰分割和值分析師。
5. 針對資料庫許可，選取資料庫許可下的描述。
6. 針對資料表和資料欄許可，選取資料表許可下的選取和描述。
7. 選擇 Grant (授予)。
8. 針對使用者 重複這些步驟DataAnalyst，其中 LF-Tag 金鑰為 Confidentiality，值為 public。

   此時，消費者帳戶中的資料分析師使用者應該能夠找到資料庫和資源連結，並透過位於 https://console.aws.amazon.com/athena/：//www.healthnet.com 的 Athena 主控台查詢共用資料表。如果沒有，請確認下列項目是否已正確設定：

   - 系統會為共用資料表建立資源連結
   - 您已授予使用者對生產者帳戶共用的 LF 標籤的存取權
   - 您授予使用者存取與資源連結和在其中建立資源連結之資料庫相關聯的 LF-Tag
   - 檢查您是否將正確的 LF 標籤指派給資源連結，以及資源連結建立所在的資料庫

步驟 4：實作具名資源方法

若要使用具名資源方法，我們會逐步引導您完成下列高階步驟：

1. 選擇性地撤銷資料庫、資料表和資料欄IAMAllowedPrincipals上的 許可。
2. 將資料許可授予取用者帳戶。
3. 接受來自 的資源共用 AWS Resource Access Manager。
4. 建立共用資料表的資源連結。
5. 將共用資料表的資料許可授予取用者。
6. 將資源連結的資料許可授予取用者。

在資料庫、資料表和資料欄**IAMAllowedPrincipals**上撤銷 的許可 （選用）

- 在本教學課程一開始，我們變更了 Lake Formation Data Catalog 設定。如果您略過該部分，則需要此步驟。如需說明，請參閱上一節中的選用步驟。

將資料許可授予取用者帳戶

1.
> **Note**
>
> 如果您以另一個使用者身分登入生產者帳戶，請先登出。

使用生產者帳戶資料湖管理員，使用 AWS 帳戶 ID、IAM 使用者名稱 （預設為 `DatalakeAdminProducer`) 和 AWS CloudFormation 堆疊建立期間指定的密碼，在 https://console.aws.amazon.com/lakeformation/：//www.healthnet.com 登入 Lake Formation 主控台。

2. 在許可頁面的資料湖許可下，選擇授予。
3. 在主體下，選擇外部帳戶，然後輸入一或多個 AWS 帳戶 IDs或 AWS 組織 IDs。如需詳細資訊，請參閱：AWS Organizations。

   生產者帳戶所屬的組織和同一 AWS 帳戶 組織內的組織會自動顯示。否則，請手動輸入帳戶 ID 或組織 ID。
4. 針對 LF 標籤或目錄資源，選擇 `Named data catalog resources`。
5. 在資料庫下，選擇資料庫 `lakeformation_tutorial_cross_account_database_named_resource`。
6. 選擇新增 LF 標籤。
7. 在資料表下，選擇所有資料表。
8. 針對資料表資料欄許可，選擇選取，然後在資料表許可下描述。
9. SelectSelect 和 Describe，在可授予許可下。
10. 或者，對於資料許可，如果需要資料欄層級許可管理，請選擇簡易資料欄型存取。
11. 選擇 Grant (授予)。

如果您尚未撤銷 的許可`IAMAllowedPrincipals`，您會收到授予許可失敗錯誤。此時，您應該會在許可、資料許可下看到透過 AWS RAM 與取用者帳戶共用的目標資料表。

從 接受資源共用 AWS RAM

> **Note**
>
> 只有以 AWS 帳戶組織為基礎的共用才需要此步驟。

1. 使用在 AWS CloudFormation 堆疊建立期間指定的 IAM 使用者名稱 （預設為 DatalakeAdminConsumer) 和密碼，使用消費者帳戶資料湖管理員登入 https：//https://console.aws.amazon.com/connect/.. 上的 AWS 主控台。
2. 在 AWS RAM 主控台的導覽窗格中，在與我共用下，資源共用下，選擇共用的 Lake Formation 資源。狀態應為待定。
3. 選擇動作和授予。
4. 確認資源詳細資訊，然後選擇接受資源共用。

   此時，消費者帳戶資料湖管理員應該能夠在 Data Catalog， Databases 下的 Lake Formation 主控台 (https://console.aws.amazon.com/lakeformation/：//) 上找到共用資源。

建立共用資料表的資源連結

- 遵循 步驟 3：使用標籤型存取控制方法實作跨帳戶共用（步驟 6) 中的指示，為共用資料表建立資源連結。命名資源連結 `amazon_reviews_table_named_resource_resource_link`。在資料庫 中建立資源連結`lakeformation_tutorial_cross_account_database_consumer`。

將共用資料表的資料許可授予取用者

若要將共用資料表的資料許可授予消費者，請完成下列步驟：

1. 在 Lake Formationconsole (https://console.aws.amazon.com/lakeformation/) 上，在許可、資料湖許可下，選擇授予。
2. 針對主體，選擇 IAM 使用者和角色，然後選擇使用者 `DataAnalyst`。
3. 針對 LF 標籤或目錄資源，選擇具名資料目錄資源。
4. 在資料庫下，選擇資料庫 `lakeformation_tutorial_cross_account_database_named_resource`。如果您在下拉式清單中看不到資料庫，請選擇載入更多。
5. 在資料表下，選擇資料表 `amazon_reviews_table_named_resource`。
6. 針對資料表和資料欄許可，選取資料表許可下的選取和描述。
7. 選擇 Grant (授予)。

將資源連結的資料許可授予取用者

除了授予資料湖使用者存取共用資料表的許可之外，您也需要授予資料湖使用者存取資源連結的許可。

1. 在 Lake Formation 主控台 (https://console.aws.amazon.com/lakeformation/) 的許可、資料湖許可下，選擇授予。
2. 針對主體，選擇 IAM 使用者和角色，然後選擇使用者 DataAnalyst。
3. 針對 LF 標籤或目錄資源，選擇具名資料目錄資源。
4. 在資料庫 下，選擇資料庫 lakeformation_tutorial_cross_account_database_consumer。如果您在下拉式清單中看不到資料庫，請選擇載入更多。
5. 在資料表下，選擇資料表 amazon_reviews_table_named_resource_resource_link。
6. 針對資源連結許可，選取資源連結許可下的描述。
7. 選擇 Grant (授予)。

   此時，消費者帳戶中的資料分析師使用者應該能夠找到資料庫和資源連結，並透過 Athena 主控台查詢共用資料表。

   如果沒有，請確認下列項目是否已正確設定：

   - 系統會為共用資料表建立資源連結
   - 您已授予使用者對生產者帳戶共用之資料表的存取權
   - 您授予使用者資源連結和建立資源連結之資料庫的存取權

步驟 5：清除 AWS 資源

若要避免您的 產生不必要的費用 AWS 帳戶，您可以刪除您用於本教學課程 AWS 的資源。

1. 使用生產者帳戶登入位於 https://console.aws.amazon.com/lakeformation/ 的 Lake Formation 主控台，並刪除或變更下列項目：

   - AWS Resource Access Manager 資源共用
   - Lake Formation 標籤
   - AWS CloudFormation 堆疊
   - Lake Formation 設定
   - AWS Glue Data Catalog
2. 使用消費者帳戶登入位於 https://console.aws.amazon.com/lakeformation/：//www.healthnet.com 的 Lake Formation 主控台，並刪除或變更下列項目：

   - Lake Formation 標籤

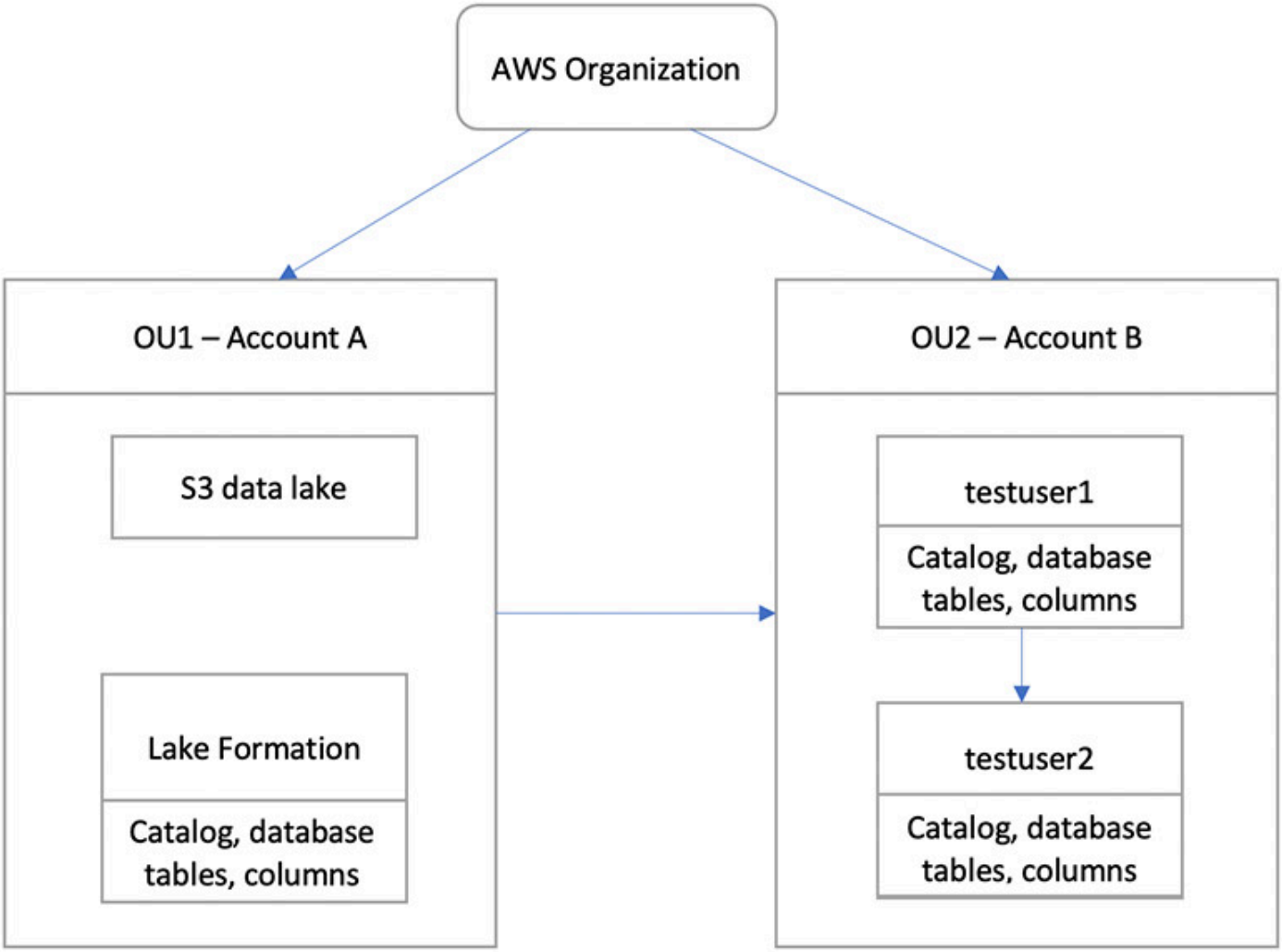
- AWS CloudFormation 堆疊

使用 Lake Formation 精細存取控制共用資料湖

本教學課程提供step-by-step說明，說明如何 AWS 帳戶 在使用 管理多個 時，使用 Lake Formation 快速輕鬆地共用資料集 AWS Organizations。您可以定義精細的許可，以控制對敏感資料的存取。

下列程序也顯示帳戶 A 的資料湖管理員如何為帳戶 B 提供精細存取，以及帳戶 B 中的使用者如何作為資料管理員，為其帳戶中的其他使用者授予對共用資料表的精細存取。每個帳戶內的資料管理員可以獨立將存取權委派給自己的使用者，為每個團隊或業務單位 (LOB) 提供自主權。

使用案例假設您使用 AWS Organizations 來管理 AWS 帳戶。一個組織單位 (OU1) 中的帳戶 A 使用者將存取權授予 OU2 中帳戶 B 的使用者。您可以在不使用 Organizations 時使用相同的方法，例如只有幾個帳戶時。下圖說明資料湖中資料集的精細存取控制。資料湖可在帳戶 A 中使用。帳戶 A 的資料湖管理員提供帳戶 B 的精細存取。圖表也顯示帳戶 B 的使用者提供帳戶 A 資料湖資料表的資料欄層級存取給帳戶 B 中的其他使用者。



主題

- 目標對象
- 先決條件
- 步驟 1：提供對另一個 帳戶的精細存取
- 步驟 2：為相同帳戶中的使用者提供精細存取

目標對象

本教學課程適用於資料管理員、資料工程師和資料分析師。下表列出本教學課程中使用的角色：

角色	描述
IAM 管理員	具有 AWS 受管政策的使用者：AdministratorAccess 。
資料湖管理員	具有 AWS 受管政策的使用者： AWSLakeFormationDataAdmin 已連接至角色。
資料分析	已AmazonAthenaFullAccess 連接 AWS 受管政策的使用者：。

先決條件

開始本教學課程之前，您必須擁有可用來以具有正確許可的管理使用者身分登入 AWS 帳戶 的 。如需詳細資訊，請參閱完成初始 AWS 組態任務。

本教學課程假設您熟悉 IAM。如需 IAM 的相關資訊，請參閱 IAM 使用者指南。

在本教學課程中，您需要下列資源：

- 兩個組織單位：
  - OU1 – 包含帳戶 A
  - OU2 – 包含帳戶 B
- 帳戶 A 中的 Amazon S3 資料湖位置 （儲存貯體）。
- 帳戶 A 中的資料湖管理員使用者。您可以使用 Lake Formation 主控台 (https://console.aws.amazon.com/lakeformation/) 或 Lake Formation API PutDataLakeSettings的操作來建立資料湖管理員。
- 帳戶 A 中設定的 Lake Formation，以及帳戶 A 中向 Lake Formation 註冊的 Amazon S3 資料湖位置。
- 帳戶 B 中的兩個使用者具有下列 IAM 受管政策：
  - testuser1 – 已AWSLakeFormationDataAdmin連接 AWS 受管政策。
  - testuser2 – 已AmazonAthenaFullAccess連接 AWS 受管政策。
- 帳戶 B Lake Formation 資料庫中的資料庫 testdb。

步驟 1：提供對另一個 帳戶的精細存取

了解帳戶 A 的資料湖管理員如何為帳戶 B 提供精細的存取。

授予對另一個 帳戶的精細存取權

1. 在帳戶 A 中的 https://console.aws.amazon.com/connect/：// AWS Management Console 登入為資料湖管理員。
2. 開啟 Lake Formation 主控台 (https://console.aws.amazon.com/lakeformation/)，然後選擇開始使用。
3. 在導覽窗格中，選擇資料庫。
4. 選擇 Create database (建立資料庫)。
5. 在資料庫詳細資訊區段中，選取資料庫。
6. 針對名稱，輸入名稱 （ 在本教學課程中，我們使用 `sampledb01`)。
7. 請確定未選取在此資料庫中僅針對新資料表使用 IAM 存取控制。離開此未選取允許我們控制 Lake Formation 的存取。
8. 選擇建立資料庫。
9. 在資料庫頁面上，選擇您的資料庫 `sampledb01`。
10. 在動作功能表中，選擇授予。
11. 在授予許可區段中，選取外部帳戶。
12. 針對 AWS 帳戶 ID 或 AWS 組織 ID，在 OU2 中輸入帳戶 B 的帳戶 ID。
13. 針對資料表，選擇您希望帳戶 B 能夠存取的資料表 （ 在此文章中，我們使用資料表 `acc_a_area`)。或者，您可以授予對 資料表中資料欄的存取權，而我們在這篇文章中執行這項操作。
14. 針對包含資料欄選擇您希望帳戶 B 有權存取的資料欄 （ 針對此文章，我們會授予 類型、名稱和識別符的許可 ）。
15. 針對資料欄，選擇包含資料欄。
16. 針對資料表許可，選取選取。
17. 針對可授予許可，選取選取。需要可授予的許可，因此帳戶 B 中的管理員使用者可以將許可授予帳戶 B 中的其他使用者。
18. 選擇 Grant (授予)。
19. 在導覽窗格中，選擇 Tables (資料表)。
20. 您可以在 AWS 帳戶 和具有存取 AWS 的組織區段中看到一個作用中的連線。

建立資源連結

Amazon Athena 等整合服務無法直接跨帳戶存取資料庫或資料表。因此，您需要建立資源連結，以便 Athena 可以存取您帳戶中的資源連結，以存取其他帳戶中的資料庫和資料表。建立資料表 (acc_a_area) 的資源連結，讓帳戶 B 使用者可以使用 Athena 查詢其資料。

1. 在帳戶 B 中，以 身分登入 AWS https：//https://console.aws.amazon.com/connect/：//testuser1https：//www.healthnet.com。
2. 在 Lake Formation 主控台 (https://console.aws.amazon.com/lakeformation/) 的導覽窗格中，選擇資料表。您應該會看到帳戶 A 提供存取權的資料表。
3. 選擇 acc_a_area 資料表。
4. 在動作功能表中，選擇建立資源連結。
5. 針對資源連結名稱，輸入名稱 （ 本教學課程為 acc_a_area_rl)。
6. 針對資料庫，選擇您的資料庫 (testdb)。
7. 選擇 Create (建立)。
8. 在導覽窗格中，選擇 Tables (資料表)。
9. 選擇 acc_b_area_rl 資料表。
10. 在動作功能表中，選擇檢視資料。

    系統會將您重新導向至 Athena 主控台，您應該會在其中看到資料庫和資料表。

    您現在可以在資料表上執行查詢，以查看從帳戶 B 為 testuser1 提供存取權的資料欄值。

步驟 2：為相同帳戶中的使用者提供精細存取

本節顯示帳戶 B (testuser1) 中的使用者如何作為資料管理員，為相同帳戶 (testuser2) 中的另一個使用者提供對共用資料表 中的資料欄名稱的精細存取aac_b_area_rl。

將精細存取權授予相同帳戶中的使用者

1. 在帳戶 B AWS 中的 https：// https://console.aws.amazon.com/connect/testuser1：// www.microsoft.com/microsoft.microsoft.com/microsoft.com/microsoft.com/
2. 在 Lake Formation 主控台的導覽窗格中，選擇資料表。

    您可以透過資料表的資源連結授予許可。若要這樣做，請在資料表頁面上，選取資源連結 acc_b_area_rl，然後在動作功能表上，選擇對目標授予。

3. 在授予許可區段中，選取我的帳戶。
4. 針對 IAM 使用者和 rolę，選擇使用者 `testuser2`。
5. 針對資料欄，選擇資料欄名稱。
6. 針對資料表許可，選取選取。
7. 選擇 Grant (授予)。

   當您建立資源連結時，只有您可以檢視和存取它。若要允許帳戶中的其他使用者存取資源連結，您需要授予資源連結本身的許可。您需要授予 DESCRIBE 或 DROP 許可。在資料表頁面上，再次選取您的資料表，然後在動作功能表上選擇授予。
8. 在授予許可區段中，選取我的帳戶。
9. 對於 IAM 使用者和角色，選取使用者 `testuser2`。
10. 針對資源連結許可選取描述。
11. 選擇 Grant (授予)。
12. 以 身分登入帳戶 B 中的 AWS 主控台`testuser2`。

    在 Athena 主控台 (https://console.aws.amazon.com/athena/) 上，您應該會看到資料庫和資料表 `acc_b_area_rl`。您現在可以在資料表上執行查詢，以查看`testuser2`可存取 的資料欄值。

加入 Lake Formation 許可

AWS Lake Formation 使用 AWS Glue Data Catalog (Data Catalog) 以目錄、資料庫和資料表的形式存放 Amazon S3 資料湖和外部資料來源的中繼資料，例如 Amazon Redshift。Data Catalog 中的中繼資料會組織在包含目錄、資料庫和資料表的三層資料階層中。它會將來自各種來源的資料組織到稱為目錄的邏輯容器。資料庫是資料表的集合。Data Catalog 也包含資源連結，這些連結是外部帳戶中共用資料庫和資料表的連結，用於跨帳戶存取資料湖中的資料。每個 AWS 帳戶每個 AWS 區域都有一個資料目錄。

Lake Formation 提供關聯式資料庫管理系統 (RDBMS) 許可模型，以授予或撤銷對 Data Catalog 中具有 Amazon S3 中基礎資料的目錄、資料庫、資料表和資料欄的存取權。

在您了解 Lake Formation 許可模型的詳細資訊之前，檢閱下列背景資訊會很有幫助：

- Lake Formation 管理的資料湖位於 Amazon Simple Storage Service (Amazon S3) 的指定位置。Data Catalog 也包含目錄物件。每個目錄代表來自 Amazon Redshift 資料倉儲、 Amazon DynamoDB 資料庫和第三方資料來源等來源的資料，例如 Snowflake、MySQL，以及透過聯合連接器整合的 30 多個外部資料來源。
- Lake Formation 會維護 Data Catalog，其中包含要匯入資料湖的來源資料中繼資料，例如日誌和關聯式資料庫中的資料，以及 Amazon S3 中資料湖中的資料。Data Catalog 也包含來自 Amazon S3 以外外部資料來源的資料中繼資料。中繼資料會組織為目錄、資料庫和資料表。中繼資料表包含結構描述、位置、分割，以及其所代表資料的其他資訊。中繼資料資料庫是資料表的集合。
- Lake Formation Data Catalog 與 使用的 Data Catalog 相同AWS Glue。您可以使用AWS Glue爬蟲程式來建立 Data Catalog 資料表，也可以使用AWS Glue擷取、轉換和載入 (ETL) 任務，在資料湖中填入基礎資料。
- Data Catalog 中的目錄、資料庫和資料表稱為 Data Catalog 資源。Data Catalog 中的資料表稱為中繼資料資料表，以區分它們與資料來源中的資料表或 Amazon S3 中的表格資料。中繼資料表在 Amazon S3 或資料來源中指向的資料稱為基礎資料。
- 委託人是使用者或角色、Amazon QuickSight 使用者或群組、透過 SAML 提供者向 Lake Formation 驗證的使用者或群組，或是跨帳戶存取控制的使用者或群組、 AWS 帳戶 ID、組織 ID 或組織單位 ID。
- AWS Glue 爬蟲程式會建立中繼資料表，但您也可以使用 Lake Formation 主控台、 API 或 AWS Command Line Interface () 手動建立中繼資料表AWS CLI。建立中繼資料表時，您必須指定位置。當您建立資料庫時，位置是選用的。資料表位置可以是 Amazon S3 位置或資料來源位置，例如 Amazon Relational Database Service (Amazon RDS) 資料庫。資料庫位置一律是 Amazon S3 位置。

- 與 Lake Formation 整合的服務，例如 Amazon Athena 和 Amazon Redshift，可以存取 Data Catalog 以取得中繼資料，並檢查執行中查詢的授權。如需整合服務的完整清單，請參閱 AWS 服務與 Lake Formation 整合。

主題

- Lake Formation 許可概觀
- Lake Formation 角色和 IAM 許可參考
- 變更資料湖的預設設定
- 隱含 Lake Formation 許可
- Lake Formation 許可參考
- 整合 IAM Identity Center
- 將 Amazon S3 位置新增至您的資料湖
- 混合存取模式
- 在 中建立物件 AWS Glue Data Catalog
- 在 Lake Formation 中使用工作流程匯入資料

Lake Formation 許可概觀

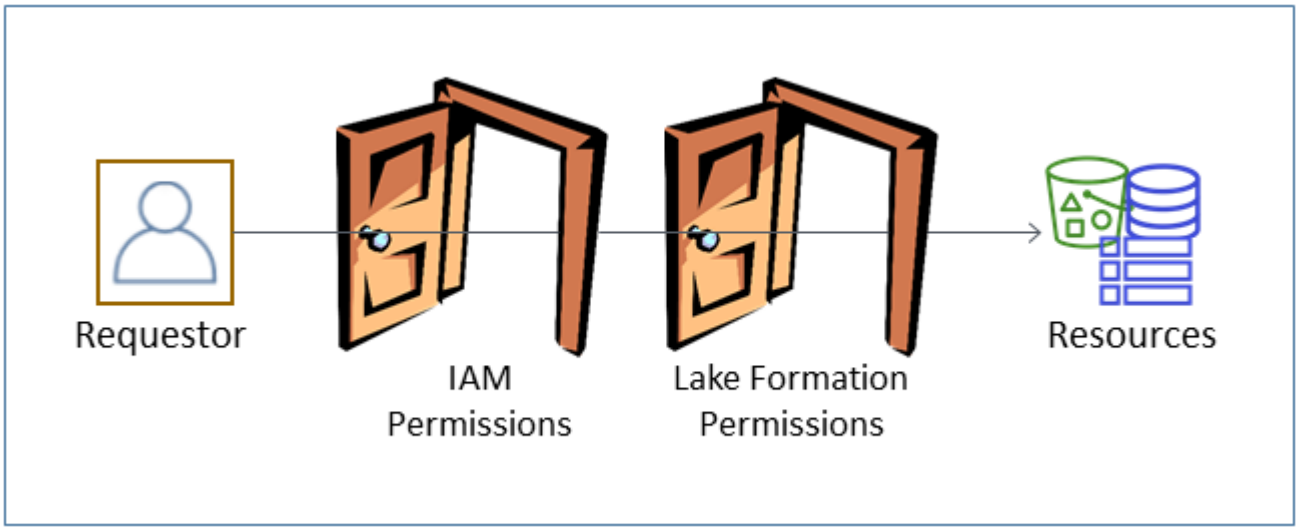
中有兩種主要類型的許可 AWS Lake Formation：

- 中繼資料存取 – Data Catalog 資源的許可 (Data Catalog 許可)。

  這些許可可讓主體在 Data Catalog 中建立、讀取、更新和刪除中繼資料資料庫和資料表。
- 基礎資料存取 – Amazon Simple Storage Service (Amazon S3) 中位置的許可 (資料存取許可和資料位置許可)。
  - 資料湖許可可讓主體讀取和寫入資料到基礎 Amazon S3 位置，資料目錄資源指向該資料。
  - 資料位置許可可讓主體建立和修改指向特定 Amazon S3 位置的中繼資料資料庫和資料表。

對於這兩個區域，Lake Formation 會使用 Lake Formation 許可和 AWS Identity and Access Management (IAM) 許可的組合。IAM 許可模型包含 IAM 政策。Lake Formation 許可模型會實作為 DBMS 樣式的 GRANT/REVOKE 命令，例如 `Grant SELECT on tableName to userName`。

當委託人提出存取 Data Catalog 資源或基礎資料的請求時，若要成功請求，必須同時通過 IAM 和 Lake Formation 的許可檢查。



Lake Formation 許可控制對 Data Catalog 資源、Amazon S3 位置和這些位置基礎資料的存取。IAM 許可控制對 Lake Formation 和 AWS Glue APIs存取。因此，雖然您可能擁有 Lake Formation 在 Data Catalog (CREATE_TABLE) 中建立中繼資料表的許可，但如果您沒有 `glue:CreateTable` API 上的 IAM 許可，您的操作會失敗。（為什麼要有`glue:`許可？ 因為 Lake Formation 使用 AWS Glue Data Catalog。）

> **Note**
>
> Lake Formation 許可僅適用於授予許可的 區域。

AWS Lake Formation 要求每個委託人（使用者或角色）都有權對 Lake Formation 管理的資源執行動作。資料湖管理員或其他具有授予 Lake Formation 許可的委託人會獲得必要的授權。

當您將 Lake Formation 許可授予委託人時，您可以選擇性地將該許可授予其他委託人。

您可以使用 Lake Formation API、 AWS Command Line Interface (AWS CLI) 或 Lake Formation 主控台的資料許可和資料位置頁面來授予和撤銷 Lake Formation 許可。

精細存取控制的方法

使用資料湖，目標是對資料進行精細存取控制。在 Lake Formation 中，這表示對 Data Catalog 資源和 Amazon S3 位置進行精細存取控制。您可以使用下列其中一種方法來實現精細存取控制。

方法	Lake Formation 許可	IAM 許可	說明
方法 1	開啟	精細分級	這是與 回溯相容性的預設方法AWS Glue。 • Open 表示將特殊許可Super授予群組 IAMAllowedPrincipals ，其中 IAMAllowedPrincipals 會自動建立，並包含 IAM 政策允許存取 Data Catalog 資源的任何 IAM 使用者和角色，而 Super許可可讓主體在授予它的資料庫或資料表上執行每個支援的 Lake Formation 操作。這實際上會導致僅由 IAM 政策控制對 Data Catalog 資源和 Amazon S3 位置的存取。如需詳細資訊，請參閱變更資料湖的預設設定及將 AWS Glue資料許可升級到 AWS Lake Formation 模型。 • 精細定義表示 IAM 政策控制對 Data Catalog 資源和個別 Amazon S3 儲存貯體的所有存取。 在 Lake Formation 主控台上，此方法會顯示為僅限 IAM 存取控制。
方法 2	精細分級	粗粒	這是建議的方法。 • 精細存取表示對 Data Catalog 資源、Amazon S3 位置和這些位置的基礎資料上的個別主體授予有限的 Lake Formation 許可。 • 粗粒化表示對個別操作和對 Amazon S3 位置的存取有更廣泛的許可。例如，粗粒 IAM 政策可能包含 "glue:*"或 "glue:Create*" 而非 "glue:CreateTables" ，讓 Lake Formation 具

方法	Lake Formation 許可	IAM 許可	說明
			有許可來控制委託人是否可以建立目錄物件。這也表示讓主體存取他們執行工作所需的 APIs，但鎖定其他 APIs 和資源。例如，您可以建立 IAM 政策，讓主體能夠建立 Data Catalog 資源並建立和執行工作流程，但無法建立AWS Glue連線或使用者定義的函數。請參閱本節稍後的範例。

Important

請注意以下事項：

- 根據預設，Lake Formation 已啟用僅使用 IAM 存取控制設定，以與現有的 AWS Glue Data Catalog 行為相容。我們建議您在轉換為使用 Lake Formation 許可後停用這些設定。如需詳細資訊，請參閱變更資料湖的預設設定。
- 資料湖管理員和資料庫建立者具有您必須了解的隱含 Lake Formation 許可。如需詳細資訊，請參閱隱含 Lake Formation 許可。

中繼資料存取控制

對於 Data Catalog 資源的存取控制，下列討論會採用具有 Lake Formation 許可的精細存取控制，以及具有 IAM 政策的粗略精細存取控制。

有兩種不同的方法來授予 Data Catalog 資源的 Lake Formation 許可：

- 具名資源存取控制 – 使用此方法，您可以透過指定資料庫或資料表名稱來授予特定資料庫或資料表的許可。授予具有此表單：

  將許可授予資源 【使用授予選項】 上的主體。

  使用授予選項，您可以允許承授者將許可授予其他委託人。
- 標籤型存取控制 – 使用此方法，您可以將一或多個 LF 標籤指派給 Data Catalog 資料庫、資料表和資料欄，並將一或多個 LF 標籤的許可授予委託人。每個 LF-Tag 都是金鑰值對，例如

department=sales。擁有符合 Data Catalog 資源上 LF 標籤的 LF 標籤的主體可以存取該資源。對於具有大量資料庫和資料表的資料湖，建議使用此方法。詳細解釋於 Lake Formation 標籤型存取控制。

委託人在資源上擁有的許可是這兩種方法所授予許可的聯集。

下表摘要說明 Data Catalog 資源上可用的 Lake Formation 許可。欄標題表示授予許可的資源。

目錄	資料庫	資料表
CREATE_DATABASE	CREATE_TABLE	ALTER
	ALTER	DROP
	DROP	DESCRIBE
	DESCRIBE	SELECT*
		INSERT*
		DELETE*

例如，在資料庫上授予 CREATE_TABLE許可。這表示允許主體在該資料庫中建立資料表。

Data Catalog 資源上會授予具有星號 (*) 的許可，但它們適用於基礎資料。例如，中繼資料資料表的 DROP 許可可讓您從資料目錄中捨棄資料表。不過，相同資料表上授予的DELETE許可可讓您使用 SQL DELETE陳述式等方式刪除 Amazon S3 中資料表的基礎資料。使用這些許可，您也可以在 Lake Formation 主控台上檢視資料表，並使用 AWS Glue API 擷取資料表的相關資訊。因此，SELECT、INSERT和 DELETE都是 Data Catalog 許可和資料存取許可。

在資料表SELECT上授予 時，您可以新增包含或排除一或多個資料欄的篩選條件。這允許對中繼資料表資料欄進行精細存取控制，限制整合服務的使用者在執行查詢時可以看到的資料欄。此功能不能僅使用 IAM 政策。

還有一個名為 的特殊許可Super。Super 許可可讓委託人在授予其資料庫或資料表上執行每個支援的 Lake Formation 操作。此許可可以與其他 Lake Formation 許可共存。例如，您可以在中繼資料資料表INSERT上授予 SELECT、 Super和 。委託人可以在資料表上執行所有支援的動作，當您撤銷時Super， SELECT和 INSERT許可仍會保留。

如需每個許可的詳細資訊，請參閱 Lake Formation 許可參考。

> **Important**
>
> 若要能夠查看另一個使用者建立的資料目錄資料表，您必須在資料表上授予至少一個 Lake Formation 許可。如果資料表上授予您至少一個許可，您也可以查看資料表的包含資料庫。

您可以使用 Lake Formation 主控台、 API 或 AWS Command Line Interface () 授予或撤銷 Data Catalog 許可AWS CLI。以下是 AWS CLI 命令的範例，可授予使用者在`retail`資料庫中建立資料表的`datalake_user1`許可。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "CREATE_TABLE" --resource '{ "Database": {"Name":"retail"}}'
```

以下是粗粒存取控制 IAM 政策的範例，該政策使用 Lake Formation 許可來補充精細存取控制。它允許對任何中繼資料資料庫或資料表進行所有操作。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "glue:*Database*",
                "glue:*Table*",
                "glue:*Partition*"
            ],
            "Resource": "*"
        }
    ]
}
```

下一個範例也是粗略粒度，但略有限制。它允許對指定帳戶和區域中 Data Catalog 中的所有中繼資料資料庫和資料表進行唯讀操作。

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

將這些政策與下列政策進行比較，該政策實作以 IAM 為基礎的精細存取控制。它僅在指定帳戶和區域中的客戶關係管理 (CRM) 中繼資料資料庫中的資料表子集上授予許可。

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

如需粗略存取控制政策的更多範例，請參閱 Lake Formation 角色和 IAM 許可參考。

基礎資料存取控制

當整合 AWS 的服務請求存取由 控制存取的 Amazon S3 位置中的資料時 AWS Lake Formation，Lake Formation 會提供臨時登入資料來存取資料。

若要讓 Lake Formation 控制對 Amazon S3 位置基礎資料的存取，請向 Lake Formation 註冊該位置。

註冊 Amazon S3 位置後，您可以開始授予下列 Lake Formation 許可：

- 指向該位置的資料目錄資料表DELETE)上的資料存取許可 SELECT(INSERT、 和 。
- 該位置上的資料位置許可。

Lake Formation 資料位置許可可控制建立指向特定 Amazon S3 位置之 Data Catalog 資源的能力。資料位置許可可為資料湖內的位置提供額外的一層安全性。當您將 CREATE_TABLE或 ALTER許可授予委託人時，您也會授予資料位置許可，以限制委託人可以建立或修改中繼資料表的位置。

Amazon S3 位置是儲存貯體下的儲存貯體或字首，但不是個別的 Amazon S3 物件。

您可以使用 Lake Formation 主控台、 API 或 ，將資料位置許可授予委託人 AWS CLI。授予的一般形式如下：

```
grant DATA_LOCATION_ACCESS to principal on S3 location [with grant option]
```

如果您包含 `with grant option`，承授者可以將許可授予其他委託人。

請記住，Lake Formation 許可一律會與 AWS Identity and Access Management (IAM) 許可搭配使用，以進行精細存取控制。對於基礎 Amazon S3 資料的讀取/寫入許可，IAM 許可的授予方式如下：

註冊位置時，您可以指定 IAM 角色，授予該位置的讀取/寫入許可。Lake Formation 在將臨時登入資料提供給整合 AWS 服務時，會擔任該角色。典型角色可能附加了下列政策，其中註冊的位置是儲存貯體 `awsexamplebucket`。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "s3:PutObject",
                "s3:GetObject",
                "s3:DeleteObject"
            ],
            "Resource": [
                "arn:aws:s3:::amzn-s3-demo-bucket/*"
            ]
```

```
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::amzn-s3-demo-bucket"
            ]
        }
    ]
}
```

Lake Formation 提供一個服務連結角色，您可以在註冊期間用來自動建立像這樣的政策。如需詳細資訊，請參閱使用 Lake Formation 的服務連結角色。

因此，註冊 Amazon S3 位置會授予該位置所需的 IAM `s3:`許可，其中許可是由用來註冊位置的角色所指定。

> **Important**
>
> 避免註冊已啟用申請者付款的 Amazon S3 儲存貯體。對於向 Lake Formation 註冊的儲存貯體，用於註冊儲存貯體的角色一律會被視為申請者。如果儲存貯體是由另一個 AWS 帳戶存取，則如果角色屬於與儲存貯體擁有者相同的帳戶，則會向儲存貯體擁有者收取資料存取費用。

對於基礎資料的讀取/寫入存取，除了 Lake Formation 許可之外，主體還需要 `lakeformation:GetDataAccess` IAM 許可。有了此許可，Lake Formation 就會授與要求存取資料所需的臨時憑證。

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

在上述政策中，您必須將資源參數設定為 '*'（全部。不支援為此許可指定任何其他資源。此組態可確保 Lake Formation 可以有效率地管理整個資料湖環境中的資料存取。

> **Note**
>
> Amazon Athena 要求使用者擁有 `lakeformation:GetDataAccess`許可。其他整合服務需要其基礎執行角色才能擁有 `lakeformation:GetDataAccess`許可。

此許可包含在 的建議政策中Lake Formation 角色和 IAM 許可參考。

總結來說，若要讓 Lake Formation 主體能夠讀取和寫入具有 Lake Formation 許可控制之存取的基礎資料：

- 向 Lake Formation 註冊包含資料的 Amazon S3 位置。
- 建立指向基礎資料位置之 Data Catalog 資料表的主體必須具有資料位置許可。
- 讀取和寫入基礎資料的委託人必須在指向基礎資料位置的資料目錄資料表上具有 Lake Formation 資料存取許可。
- 當基礎資料位置向 Lake Formation `lakeformation:GetDataAccess` 註冊時，讀取和寫入基礎資料的委託人必須具有 IAM 許可。

> **Note**
>
> 如果您可以透過 IAM 或 Amazon S3 政策存取 Amazon S3 API 或主控台，Lake Formation 許可模型不會阻止透過 Amazon S3 API 或主控台存取 Amazon S3 位置。您可以將 IAM 政策連接至主體以封鎖此存取。

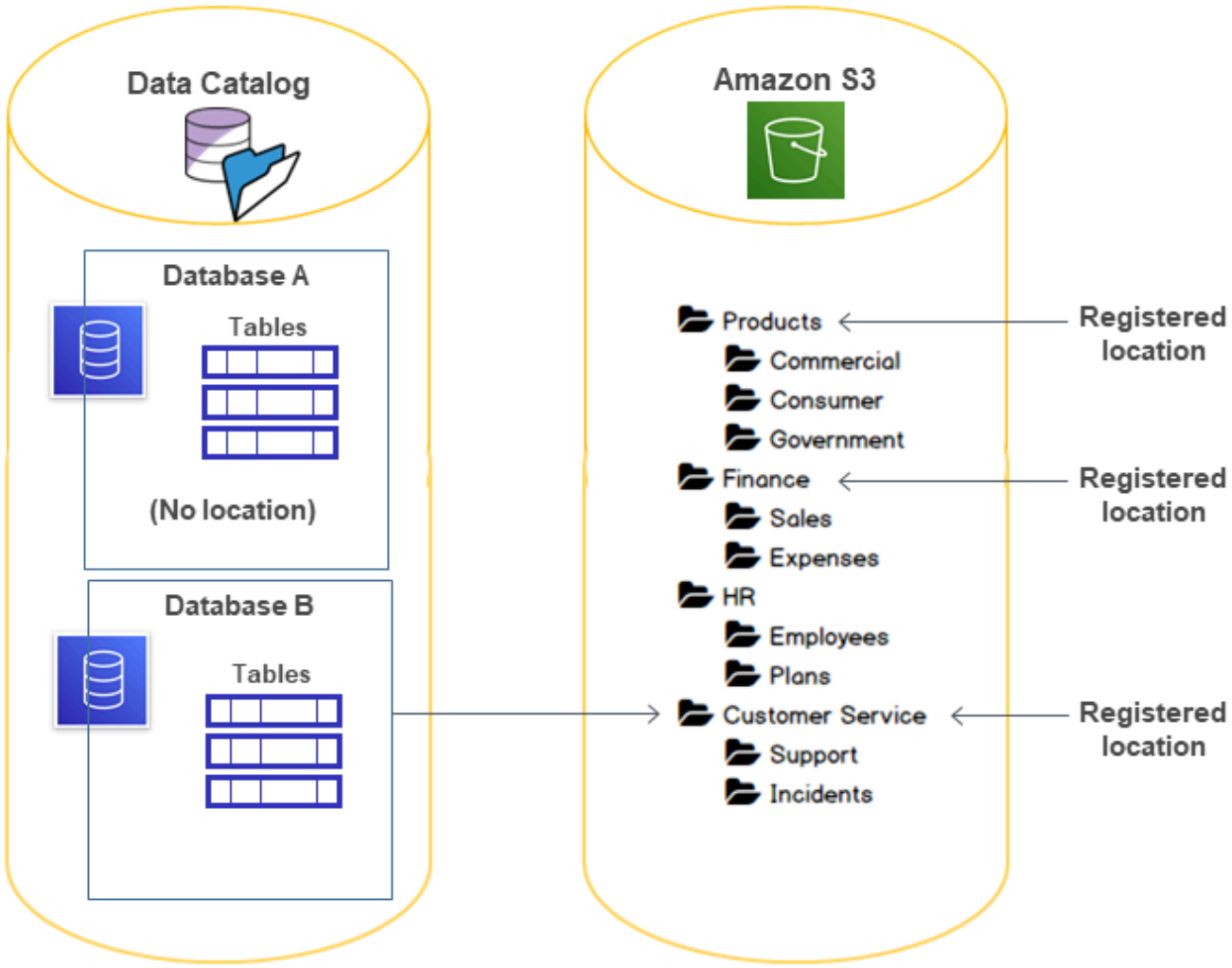
有關資料位置許可的詳細資訊

資料位置許可會管理 Data Catalog 資料庫和資料表上建立和更新操作的結果。規則如下：

- 委託人必須在 Amazon S3 位置上具有明確或隱含的資料位置許可，才能建立或更新指定該位置的資料庫或資料表。
- 明確許可`DATA_LOCATION_ACCESS`是使用 主控台、API 或 授予 AWS CLI。
- 當資料庫具有指向已註冊位置的位置屬性、主體具有資料庫的`CREATE_TABLE`許可，且主體嘗試在該位置或子位置建立資料表時，就會授予隱含許可。

- 如果將資料位置許可授予委託人，則委託人在所有子位置都有資料位置許可。
- 主體不需要資料位置許可，即可對基礎資料執行讀取/寫入操作。具備 SELECT或 INSERT資料存取許可就已足夠。資料位置許可僅適用於建立指向位置的 Data Catalog 資源。

考慮下圖中顯示的案例。



在這張圖中：

- Amazon S3 儲存貯體 Products、 Finance和 Customer Service 已向 Lake Formation 註冊。
- Database A 沒有位置屬性，且Database B具有指向儲存Customer Service貯體的位置屬性。
- 使用者在兩個資料庫CREATE_TABLE上datalake_user都有 。

- 使用者僅datalake_user獲得儲存Products貯體上的資料位置許可。

以下是使用者datalake_user嘗試在特定位置的特定資料庫中建立目錄資料表時的結果。

datalake_user 嘗試建立資料表的位置

資料庫和位置	成功或失敗	原因
資料庫 A 位於 Finance/Sales	失敗	沒有資料位置許可
資料庫 A 位於 Products	成功	具有資料位置許可
資料庫 A 位於 HR/Plans	成功	位置未註冊
資料庫 B 位於 Customer Service/Incidents	成功	資料庫的位置屬性位於 Customer Service

如需詳細資訊，請參閱下列內容：

- 將 Amazon S3 位置新增至您的資料湖
- Lake Formation 許可參考
- Lake Formation 角色和 IAM 許可參考

Lake Formation 角色和 IAM 許可參考

本節列出一些建議的 Lake Formation 角色及其建議的 AWS Identity and Access Management (IAM) 許可。如需 Lake Formation 許可的詳細資訊，請參閱 the section called “Lake Formation 許可參考”。

AWS Lake Formation 角色

下表列出建議 AWS Lake Formation 的角色。

Lake Formation 角色

人物	描述
IAM 管理員（超級使用者）	（必要）可建立 IAM 使用者和角色的使用者。具有 AdministratorAccess AWS 受管政策。具有所有 Lake Formation 資源

人物	描述
	的所有許可。可以新增資料湖管理員。如果也未指定資料湖管理員，則無法授予 Lake Formation 許可。
資料湖管理員	（必要）可註冊 Amazon S3 位置、存取 Data Catalog、建立資料庫、建立和執行工作流程、將 Lake Formation 許可授予其他使用者，以及檢視 AWS CloudTrail 日誌的使用者。IAM 許可少於 IAM 管理員，但足以管理資料湖。無法新增其他資料湖管理員。
唯讀管理員	（選用）可檢視主體、Data Catalog 資源、許可和 AWS CloudTrail 日誌的使用者，無需進行更新的許可。
資料工程師	（選用）可建立資料庫、建立和執行爬蟲程式和工作流程，以及對爬蟲程式和工作流程建立的資料目錄資料表授予 Lake Formation 許可的使用者。建議您讓所有資料工程師建立資料庫。如需詳細資訊，請參閱建立資料庫。
資料分析	（選用）可以使用 對資料湖執行查詢的使用者，例如 Amazon Athena。僅具有執行查詢的足夠許可。
工作流程角色	（必要）代表使用者執行工作流程的角色。當您從藍圖建立工作流程時，您可以指定此角色。

> **Note**
>
> 在 Lake Formation 中，資料庫建立後新增的資料湖管理員可以授予許可，但不會自動擁有資料存取許可，例如 SELECT 或 DESCRIBE。建立資料庫的管理員會收到這些資料庫的SUPER許可。此行為是刻意的，雖然所有管理員都可以授予自己必要的許可，但這些許可不會自動套用到預先存在的資源。因此，管理員必須明確授予自己對資料庫的存取權，然後才能獲得管理員權限。

AWS Lake Formation 的 受管政策

您可以使用 AWS 受管政策和內嵌政策 AWS Lake Formation ，授予使用 所需的 AWS Identity and Access Management (IAM) 許可。下列 AWS 受管政策適用於 Lake Formation。

AWS 受管政策：AWSLakeFormationDataAdmin

AWSLakeFormationDataAdmin 政策會授予 AWS Lake Formation 和相關服務的管理存取權，例如 AWS Glue 來管理資料湖。

您可以AWSLakeFormationDataAdmin連接到您的使用者、群組和角色。

許可詳細資訊

- CloudTrail – 允許主體檢視 AWS CloudTrail 日誌。這是檢閱資料湖設定中的任何錯誤所必需的。
- Glue – 允許主體檢視、建立和更新 Data Catalog 中的中繼資料表和資料庫。這包括以 Get、List、、Delete、 Create Update和 開頭的 API 操作Search。這是管理資料湖資料表中繼資料的必要項目。
- IAM – 允許主體擷取連接到角色的 IAM 使用者、角色和政策的相關資訊。這是資料管理員檢閱和列出 IAM 使用者和角色以授予 Lake Formation 許可的必要項目。
- Lake Formation – 授予資料湖管理員管理資料湖所需的 Lake Formation 許可。
- S3 – 允許主體擷取 Amazon S3 儲存貯體及其位置的相關資訊，以設定資料湖的資料位置。

```
"Statement": [
        {
            "Sid": "AWSLakeFormationDataAdminAllow",
            "Effect": "Allow",
            "Action": [
                "lakeformation:*",
                "cloudtrail:DescribeTrails",
                "cloudtrail:LookupEvents",
                "glue:CreateCatalog",
  "glue:UpdateCatalog",
                "glue:DeleteCatalog",
  "glue:GetCatalog",
         "glue:GetCatalogs",
                "glue:GetDatabase",
                "glue:GetDatabases",
                "glue:CreateDatabase",
                "glue:UpdateDatabase",
                "glue:DeleteDatabase",
                "glue:GetConnections",
                "glue:SearchTables",
                "glue:GetTable",
                "glue:CreateTable",
```

```
                "glue:UpdateTable",
                "glue:DeleteTable",
                "glue:GetTableVersions",
                "glue:GetPartitions",
                "glue:GetTables",
                "glue:ListWorkflows",
                "glue:BatchGetWorkflows",
                "glue:DeleteWorkflow",
                "glue:GetWorkflowRuns",
                "glue:StartWorkflowRun",
                "glue:GetWorkflow",
                "s3:ListBucket",
                "s3:GetBucketLocation",
                "s3:ListAllMyBuckets",
                "s3:GetBucketAcl",
                "iam:ListUsers",
                "iam:ListRoles",
                "iam:GetRole",
                "iam:GetRolePolicy"
            ],
            "Resource": "*"
        },
        {
            "Sid": "AWSLakeFormationDataAdminDeny",
            "Effect": "Deny",
            "Action": [
                "lakeformation:PutDataLakeSettings"
            ],
                "Resource": "*"
        }
    ]
}
```

> **Note**
>
> 此AWSLakeFormationDataAdmin政策不會授予資料湖管理員所需的所有許可。建立和執行工作流程，並使用服務連結角色 註冊位置時，需要額外的許可AWSServiceRoleForLakeFormationDataAccess。如需詳細資訊，請參閱建立資料湖管理員及使用 Lake Formation 的服務連結角色。

AWS 受管政策：AWSLakeFormationCrossAccountManager

AWSLakeFormationCrossAccountManager 政策透過 Lake Formation 提供對 AWS Glue 資源的跨帳戶存取權，並授予對 AWS Organizations 和 等其他必要服務的讀取存取權 AWS RAM。

您可以AWSLakeFormationCrossAccountManager連接到您的使用者、群組和角色。

許可詳細資訊

此政策包含以下許可。

- Glue – 允許主體設定或刪除 Data Catalog 資源政策以進行存取控制。
- Organizations – 允許主體擷取組織的帳戶和組織單位 (OU) 資訊。
- ram:CreateResourceShare – 允許主體建立資源共享。
- ram:UpdateResourceShare –允許主體修改指定資源共享的某些屬性。
- ram:DeleteResourceShare – 允許主體刪除指定的資源共享。
- ram:AssociateResourceShare – 允許主體將指定的主體清單和資源清單新增至資源共享。
- ram:DisassociateResourceShare – 允許主體從參與指定的資源共享中移除指定的主體或資源。
- ram:GetResourceShares– 允許主體擷取您擁有或與您共用之資源共用的詳細資訊。
- ram:RequestedResourceType – 允許主體擷取資源類型（資料庫、資料表或目錄）。
- AssociateResourceSharePermission – 允許主體新增或取代資源共享中包含的資源類型的 AWS RAM 許可。您可以在資源共享中，只擁有一個與每個資源類型相關聯的許可。

```
{
    "Version": "2012-10-17",
    "Statement": [{
            "Sid": "AllowCreateResourceShare",
            "Effect": "Allow",
            "Action": [
                "ram:CreateResourceShare"
            ],
            "Resource": "*",
            "Condition": {
                "StringLikeIfExists": {
                    "ram:RequestedResourceType": [
                        "glue:Table",
                        "glue:Database",
```

```
                    "glue:Catalog"
                ]
            }
        }
    },
    {
        "Sid": "AllowManageResourceShare",
        "Effect": "Allow",
        "Action": [
            "ram:UpdateResourceShare",
            "ram:DeleteResourceShare",
            "ram:AssociateResourceShare",
            "ram:DisassociateResourceShare",
            "ram:GetResourceShares"
        ],
        "Resource": "*",
        "Condition": {
            "StringLike": {
                "ram:ResourceShareName": [
                    "LakeFormation*"
                ]
            }
        }
    },
    {
        "Sid": "AllowManageResourceSharePermissions",
        "Effect": "Allow",
        "Action": [
            "ram:AssociateResourceSharePermission"
        ],
        "Resource": "*",
        "Condition": {
            "ArnLike": {
                "ram:PermissionArn": [
                    "arn:aws:ram::aws:permission/AWSRAMLFEnabled*"
                ]
            }
        }
    },
    {
        "Sid": "AllowXAcctManagerPermissions",
        "Effect": "Allow",
        "Action": [
            "glue:PutResourcePolicy",
```

```
                "glue:DeleteResourcePolicy",
                "organizations:DescribeOrganization",
                "organizations:DescribeAccount",
                "ram:Get*",
                "ram:List*"
            ],
            "Resource": "*"
        },
        {
            "Sid": "AllowOrganizationsPermissions",
            "Effect": "Allow",
            "Action": [
                "organizations:ListRoots",
                "organizations:ListAccountsForParent",
                "organizations:ListOrganizationalUnitsForParent"
            ],
            "Resource": "*"
        }
    ]
}
```

AWS 受管政策：AWSGlueConsoleFullAccess

AWSGlueConsoleFullAccess 政策會授予 AWS Glue 資源的完整存取權 AWS Management Console。如果您依照此政策中指定的資源命名慣例，使用者就能擁有完整的主控台功能。此政策通常連接到 AWS Glue 主控台的使用者。

此外，AWS GlueLake Formation 會擔任服務角色`AWSGlueServiceRole`，以允許存取相關服務，包括 Amazon Elastic Compute Cloud (Amazon EC2)、Amazon Simple Storage Service (Amazon S3) 和 Amazon CloudWatch。

AWS managed policy:LakeFormationDataAccessServiceRolePolicy

此政策會連接到名為 的服務連結角色`ServiceRoleForLakeFormationDataAccess`，該角色允許服務在您請求時對資源執行動作。您無法將此政策連接至您的 IAM 身分。

此政策允許 Lake Formation 整合 AWS 服務，例如 Amazon Athena 或 Amazon Redshift，使用服務連結角色來探索 Amazon S3 資源。

如需詳細資訊，請參閱 使用 Lake Formation 的服務連結角色。

許可詳細資訊

此政策包含以下許可。

- `s3:ListAllMyBuckets` – 傳回請求已驗證寄件者擁有的所有儲存貯體清單。

```
{
 "Version": "2012-10-17",
 "Statement": [
  {
   "Sid": "LakeFormationDataAccessServiceRolePolicy",
   "Effect": "Allow",
   "Action": [
    "s3:ListAllMyBuckets"
   ],
   "Resource": [
    "arn:aws:s3:::*"
   ]
  }
 ]
}
```

AWS 受管政策的 Lake Formation 更新

檢視自此服務開始追蹤這些變更以來，Lake Formation AWS 受管政策更新的詳細資訊。

變更	描述	日期
Lake Formation 已更新`AWSLakeFormationCrossAccountManager` 政策。	Lake Formation 透過將`StringLike` 條件運算子取代為允許 IAM 執行 ARN 格式檢查的運算`ArnLike`子，增強了 AWSLakeFormationCrossAccountManager 政策。	2025 年 1 月
Lake Formation 已更新`AWSLakeFormationDataAdmin` 政策。	Lake Formation 新增下列 AWS Glue Data Catalog CRUD APIs做為多目錄功能的一部分，以增強 AWSLakeFormationDataAdmin 政策。 • glue：CreateCatalog • glue：UpdateCatalog	2024 年 12 月

變更	描述	日期
	• glue：DeleteCatalog • glue：GetCatalog • glue：GetCatalogs 此受管政策變更旨在確保 Lake Formation 管理員角色預設具有這些新操作的 IAM 許可。	
Lake Formation 已更新`AWSLakeFormationCrossAccountManager` 政策。	Lake Formation 透過將 Sid 元素新增至政策陳述式來增強 AWSLakeFormationCrossAccountManager 政策。	2024 年 3 月
Lake Formation 已更新`AWSLakeFormationDataAdmin` 政策。	Lake Formation 透過將 Sid 元素新增至政策陳述式並移除備援動作來增強 AWSLakeFormationDataAdmin 政策。	2024 年 3 月
Lake Formation 已更新`LakeFormationDataAccessServiceRolePolicy` 政策。	Lake Formation 透過將 Sid 元素新增至政策陳述式來增強 LakeFormationDataAccessServiceRolePolicy 政策。	2024 年 2 月
Lake Formation 已更新`AWSLakeFormationCrossAccountManager` 政策。	Lake Formation 透過新增在混合存取模式中啟用跨帳戶資料共用的新許可，增強了 AWSLakeFormationCrossAccountManager 政策。	2023 年 10 月

變更	描述	日期
Lake Formation 已更新`AWSLakeFormationCrossAccountManager` 政策。	Lake Formation 增強了 AWSLakeFormationCrossAccountManager 政策，在第一次共用資源時，每個收件人帳戶只建立一個資源共用。之後與相同帳戶共用的所有資源都會連接到相同的資源共用。	2022 年 5 月 6 日
Lake Formation 開始追蹤變更。	Lake Formation 開始追蹤其 AWS 受管政策的變更。	2022 年 5 月 6 日

角色建議的許可

以下是每個角色的建議許可。IAM 管理員不包含在內，因為該使用者具有所有資源的所有許可。

主題

- Data Lake 管理員許可
- 唯讀管理員許可
- 資料工程師許可
- 資料分析師許可
- 工作流程角色許可

Data Lake 管理員許可

> **Important**
>
> 在下列政策中，將 *<account-id>* 取代為有效的 AWS 帳號，並將 *<workflow_role>* 取代為具有執行工作流程許可的角色名稱，如 中所定義工作流程角色許可。

政策類型	政策
AWS 受管政策	• `AWSLakeFormationDataAdmin`

政策類型	政策
	• LakeFormationDataAccessServiceRolePolicy（服務連結角色政策） • AWSGlueConsoleFullAccess (選用) • CloudWatchLogsReadOnlyAccess (選用) • AWSLakeFormationCrossAccountManager (選用) • AmazonAthenaFullAccess (選用) 如需選用 AWS 受管政策的相關資訊，請參閱 the section called "建立資料湖管理員"。

內嵌政策（用於建立 Lake Formation 服務連結角色）	{
	"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"
	}
	]
	}

政策類型	政策

（選用）內嵌政策（工作流程角色的通行政策）。只有在資料湖管理員建立和執行工作流程時，才需要這樣做。	{
	"Version": "2012-10-17",
	"Statement": [
	{
	"Sid": "PassRolePermissions",
	"Effect": "Allow",
	"Action": [
	"iam:PassRole"
	],
	"Resource": [
	"arn:aws:iam:: *<account-id>* :role/*<workflow_role>* "
	]
	}
	]
	}

（選用）內嵌政策（如果您的帳戶授予或接收跨帳戶 Lake Formation 許可）。此政策用於接受或拒絕 AWS RAM 資源共享邀請，以及啟用將跨帳戶許可授予組織。 `ram:EnableSharingWithAwsOrganization` 僅適用於管理帳戶中的資料湖管理員 AWS Organizations 。	{
	"Version": "2012-10-17",
	"Statement": [
	{
	"Effect": "Allow",
	"Action": [
	"ram:AcceptResourceShareInvitation",
	"ram:RejectResourceShareInvitation",
	"ec2:DescribeAvailabilityZones",
	"ram:EnableSharingWithAwsOrganization"
	],
	"Resource": "*"
	}
	]
	}

唯讀管理員許可

Policy type (政策類型)	政策
內嵌政策（基本）	{ "Version":"2012-10-17", "Statement":[{ "Effect":"Allow", "Action":["lakeformation:GetEffectivePermissionsForPath", "lakeformation:ListPermissions", "lakeformation:ListDataCellsFilter", "lakeformation:GetDataCellsFilter", "lakeformation:SearchDatabasesByLFTags", "lakeformation:SearchTablesByLFTags", "lakeformation:GetLFTag", "lakeformation:ListLFTags", "lakeformation:GetResourceLFTags", "lakeformation:ListLakeFormationOptions", "cloudtrail:DescribeTrails", "cloudtrail:LookupEvents", "glue:GetDatabase", "glue:GetDatabases", "glue:GetConnections", "glue:SearchTables", "glue:GetTable", "glue:GetTableVersions", "glue:GetPartitions", "glue:GetTables", "glue:GetWorkflow", "glue:ListWorkflows", "glue:BatchGetWorkflows", "glue:GetWorkflowRuns", "glue:GetWorkflow", "s3:ListBucket", "s3:GetBucketLocation", "s3:ListAllMyBuckets", "s3:GetBucketAcl", "iam:ListUsers",

Policy type (政策類型)	政策

	"iam:ListRoles",
	"iam:GetRole",
	"iam:GetRolePolicy"
	],
	"Resource":"*"
	},
	{
	"Effect":"Deny",
	"Action":[
	"lakeformation:PutDataLakeSettings"
	],
	"Resource":"*"
	}
	]
	}

資料工程師許可

> **Important**
>
> 在下列政策中，將 *<account-id>* 取代為有效的 AWS 帳號，並將 *<workflow_role>* 取代為工作流程角色的名稱。

政策類型	政策
AWS 受管政策	AWSGlueConsoleFullAccess
內嵌政策（基本）	{ "Version": "2012-10-17", "Statement": [{ "Effect": "Allow", "Action": ["lakeformation:GetDataAccess", "lakeformation:GrantPermissions", "lakeformation:RevokePermissions", "lakeformation:BatchGrantPermissions",

政策類型	政策
	"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": "*" }] }

政策類型	政策

內嵌政策（適用於受管資料表上的操作，包括交易中的操作）	{
	"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": "*"
	}
	]
	}

政策類型	政策

內嵌政策（適用於使用 Lake Formation 標籤型存取控制 (LF-TBAC) 方法的中繼資料存取控制）	{
	"Version": "2012-10-17",
	"Statement": [
	{
	"Effect": "Allow",
	"Action": [
	"lakeformation:AddLFTagsToResource",
	"lakeformation:RemoveLFTagsFromResou
	rce",
	"lakeformation:GetResourceLFTags",
	"lakeformation:ListLFTags",
	"lakeformation:GetLFTag",
	"lakeformation:SearchTablesByLFTags",
	"lakeformation:SearchDatabasesByLFTags"
	],
	"Resource": "*"
	}
	]
	}

內嵌政策（工作流程角色的通行政策）	{
	"Version": "2012-10-17",
	"Statement": [
	{
	"Sid": "PassRolePermissions",
	"Effect": "Allow",
	"Action": [
	"iam:PassRole"
	],
	"Resource": [
	"arn:aws:iam:: *<account-id>* :role/*<workflow
	_role>* "
	]
	}
	]
	}

資料分析師許可

政策類型	政策
AWS 受管政策	AmazonAthenaFullAccess

內嵌政策（基本）	{
	"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": "*"
	}
	]
	}

（選用）內嵌政策（適用於受管資料表的操作，包括交易中的操作）	{ "Version": "2012-10-17", "Statement": [{ "Effect": "Allow", "Action": ["lakeformation:StartTransaction", "lakeformation:CommitTransaction", "lakeformation:CancelTransaction", "lakeformation:ExtendTransaction", "lakeformation:DescribeTransaction",

政策類型	政策
	"lakeformation:ListTransactions", "lakeformation:GetTableObjects", "lakeformation:UpdateTableObjects", "lakeformation:DeleteObjectsOnCancel"], "Resource": "*" }] }

工作流程角色許可

此角色具有執行工作流程所需的許可。您可以在建立工作流程時指定具有這些許可的角色。

> **Important**
>
> 在下列政策中，將 *<region>* 取代為有效的 AWS 區域識別符（例如 us-east-1)、將 *<account-id>* 取代為有效的 AWS 帳號、將 *<workflow_role>* 取代為工作流程角色的名稱，並將 *<your-s3-cloudtrail-bucket>* 取代為 AWS CloudTrail 日誌的 Amazon S3 路徑。

政策類型	政策
AWS 受管政策	AWSGlueServiceRole
內嵌政策（資料存取）	{ "Version": "2012-10-17", "Statement": [{ "Sid": "Lakeformation", "Effect": "Allow", "Action": ["lakeformation:GetDataAccess", "lakeformation:GrantPermissions"], "Resource": "*"

政策類型	政策
	<pre> }] }</pre>

內嵌政策（工作流程角色的通行政策）	<pre>{
	    "Version": "2012-10-17",
	    "Statement": [
	        {
	            "Sid": "PassRolePermissions",
	            "Effect": "Allow",
	            "Action": [
	                "iam:PassRole"
	            ],
	            "Resource": [
	                "arn:aws:iam:: <account-id> :role/<workflow_role> "
	            ]
	        }
	    ]
	}</pre>

內嵌政策（用於在資料湖外擷取資料，例如 AWS CloudTrail 日誌）	<pre>{
	    "Version": "2012-10-17",
	    "Statement": [
	        {
	            "Effect": "Allow",
	            "Action": ["s3:GetObject", "s3:ListBucket"],
	            "Resource": ["arn:aws:s3::: <your-s3-cloudtrail-bucket> /*"]
	        }
	    ]
	}</pre>

變更資料湖的預設設定

為了維持與 的回溯相容性AWS Glue， AWS Lake Formation 具有下列初始安全設定：

- Super 許可會授予所有現有 AWS Glue Data Catalog 資源IAMAllowedPrincipals上的 群組。

- 新的 Data Catalog 資源已啟用「僅使用 IAM 存取控制」設定。

這些設定可有效導致僅由 AWS Identity and Access Management (IAM) 政策控制對 Data Catalog 資源和 Amazon S3 位置的存取。個別 Lake Formation 許可未生效。

`IAMAllowedPrincipals` 群組包含 IAM 政策允許存取 Data Catalog 資源的任何 IAM 使用者和角色。`Super` 許可可讓委託人在授予其資料庫或資料表上執行每個支援的 Lake Formation 操作。

若要變更安全設定，以便由 Lake Formation 許可管理對 Data Catalog 資源 （ 資料庫和資料表 ） 的存取，請執行下列動作：

1. 變更新資源的預設安全設定。如需說明，請參閱 變更預設許可模型或使用混合存取模式。
2. 變更現有 Data Catalog 資源的設定。如需說明，請參閱 將AWS Glue資料許可升級到 AWS Lake Formation 模型。

使用 Lake Formation **`PutDataLakeSettings`** API 操作變更預設安全設定

您也可以使用 Lake Formation PutDataLakeSettings API 操作來變更預設安全設定。此動作會做為引數，選用的目錄 ID 和 DataLakeSettings 結構。

若要在新資料庫和資料表上強制執行 Lake Formation 的中繼資料和基礎資料存取控制，請如下編寫`DataLakeSettings`結構的程式碼。

> **Note**
>
> 將 *<AccountID>* 取代為有效的 AWS 帳戶 ID，並將 *<Username>* 取代為有效的 IAM 使用者名稱。您可以將多個使用者指定為資料湖管理員。

```
{
    "DataLakeSettings": {
        "DataLakeAdmins": [
            {
                "DataLakePrincipalIdentifier":
 "arn:aws:iam::<AccountId>:user/<Username>"
            }
        ],
        "CreateDatabaseDefaultPermissions": [],
        "CreateTableDefaultPermissions": []
    }
```

```
}
```

您也可以將結構編碼如下。省略 CreateDatabaseDefaultPermissions或 CreateTableDefaultPermissions 參數等同於傳遞空清單。

```
{
    "DataLakeSettings": {
        "DataLakeAdmins": [
            {
                "DataLakePrincipalIdentifier":
 "arn:aws:iam::<AccountId>:user/<Username>"
            }
        ]
    }
}
```

此動作會有效地撤銷新資料庫和資料表上IAMAllowedPrincipals群組的所有 Lake Formation 許可。建立資料庫時，您可以覆寫此設定。

若要僅在新資料庫和資料表上由 IAM 強制執行中繼資料和基礎資料存取控制，請依下列方式對DataLakeSettings結構編寫程式碼。

```
{
    "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"
                ]
            }
        ]
    }
}
```

這會將 Super Lake Formation 許可授予新資料庫和資料表上的 IAMAllowedPrincipals群組。建立資料庫時，您可以覆寫此設定。

> **Note**
>
> 在上述DataLakeSettings結構中， 的唯一允許值DataLakePrincipalIdentifier為 IAM_ALLOWED_PRINCIPALS，而 的唯一允許值Permissions為 ALL。

隱含 Lake Formation 許可

AWS Lake Formation 授予資料湖管理員、資料庫建立者和資料表建立者的下列隱含許可。

資料湖管理員

- 可以Describe存取資料目錄中的所有資源，但直接從另一個帳戶共用給不同委託人的資源除外。管理員無法撤銷此存取權。
- 擁有資料湖中任何地方的資料位置許可。
- 可以將 Data Catalog 中任何資源的存取權授予或撤銷給任何委託人（包括自己）。管理員無法撤銷此存取權。
- 可以在 Data Catalog 中建立資料庫。
- 可以將建立資料庫的許可授予其他使用者。

> **Note**
>
> 只有在資料湖管理員具有執行此操作的 IAM 許可時，才能註冊 Amazon S3 位置。本指南中建議的資料湖管理員政策會授予這些許可。此外，資料湖管理員沒有隱含許可來捨棄由其他人建立的資料庫或更改/捨棄資料表。不過，他們可以授予自己許可。

如需資料湖管理員的詳細資訊，請參閱 建立資料湖管理員。

目錄建立者

- 擁有其所建立目錄的所有目錄許可、擁有其在目錄中建立之資料庫和資料表的許可，並可以授予相同 AWS 帳戶中的其他主體在目錄中建立資料庫和資料表的許可。同時具有 AWSLakeFormationCrossAccountManager AWS 受管政策的目錄建立者可以將目錄的許可授予其他 AWS 帳戶或組織。

  資料湖管理員可以使用 Lake Formation 主控台或 API 來指定目錄建立者。

  > **Note**
  >
  > 目錄建立者不會隱含地對其他人在目錄中建立的資料庫和資料表擁有許可。

如需建立目錄的詳細資訊，請參閱 將您的資料帶入 AWS Glue Data Catalog。

資料庫建立者

- 擁有其所建立資料庫的所有資料庫許可、擁有其在資料庫中建立之資料表的許可，並可以授予相同 AWS 帳戶中的其他主體在資料庫中建立資料表的許可。同時具有 AWSLakeFormationCrossAccountManager AWS 受管政策的資料庫建立者可以將資料庫的許可授予其他 AWS 帳戶或組織。

  資料湖管理員可以使用 Lake Formation 主控台或 API 來指定資料庫建立者。

  > **Note**
  >
  > 資料庫建立者不會隱含地擁有其他人在資料庫中建立之資料表的許可。

如需詳細資訊，請參閱建立資料庫。

資料表建立者

- 對其建立的資料表擁有所有許可。
- 可以將他們建立的所有資料表的許可授予相同 AWS 帳戶中的主體。
- 如果他們擁有 AWSLakeFormationCrossAccountManager AWS 受管政策，則可以將他們建立的所有資料表的許可授予其他 AWS 帳戶或組織。
- 可以檢視包含其所建立資料表的資料庫。

Lake Formation 許可參考

若要執行 AWS Lake Formation 操作，主體需要 Lake Formation 許可和 AWS Identity and Access Management (IAM) 許可。您通常會使用粗粒存取控制政策授予 IAM 許可，如 中所述the section called “Lake Formation 許可概觀 ”。您可以使用 主控台、 API 或 AWS Command Line Interface () 授予 Lake Formation 許可AWS CLI。

若要了解如何授予或撤銷 Lake Formation 許可，請參閱 the section called “授予資料許可”和 the section called “授予資料位置許可”。

> **Note**
>
> 本節中的範例示範如何將許可授予相同 AWS 帳戶中的主體。如需跨帳戶授與的範例，請參閱 the section called “跨帳戶資料共用”。

每個資源類型的 Lake Formation 許可

以下是每種資源類型可用的有效 Lake Formation 許可：

資源	權限
Catalog	ALL (Super)、超級使用者
	ALTER
	CREATE_DATABASE
	DESCRIBE
	DROP
Database	ALL (Super)
	ALTER
	CREATE_TABLE
	DESCRIBE

資源	權限
	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
LF-Tag values	ASSOCIATE
	DESCRIBE
	GrantWithLFTagExpression

資源	權限
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
Table with column filter	SELECT

主題

- Lake Formation 授予和撤銷 AWS CLI 命令

- Lake Formation 許可

Lake Formation 授予和撤銷 AWS CLI 命令

本節中的每個許可描述都包含使用 AWS CLI 命令授予許可的範例。以下是 Lake Formation grant-permissions和 revoke-permissions AWS CLI 命令的摘要。

```
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>]
```

如需這些命令的詳細說明，請參閱AWS CLI 《 命令參考》中的 grant-permissions 和 revoke-permissions。本節提供 `--principal`選項的其他資訊。

`--principal` 選項的值為下列其中一項：

- (IAM) 使用者或角色的 Amazon Resource Name AWS Identity and Access Management (ARN)
- 透過 SAML 提供者進行身分驗證的使用者或群組 ARN，例如 Microsoft Active Directory Federation Service (AD FS)
- Amazon QuickSight 使用者或群組的 ARN
- 對於跨帳戶許可、 AWS 帳戶 ID、組織 ID 或組織單位 ID
- 對於 IAM Identity Center 使用者或群組，IAM Identity Center 使用者或群組 ARN。

以下是所有`--principal`類型的語法和範例。

Principal 是 IAM 使用者

語法:

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/<user-name>
```

範例：

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1
```

Principal 是 IAM 角色

語法:

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:role/<role-name>
```

範例：

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:role/workflowrole
```

委託人是透過 SAML 提供者進行身分驗證的使用者

語法：

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:saml-
provider/<SAMLproviderName>:user/<user-name>
```

範例：

```
--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
```

委託人是透過 SAML 提供者進行驗證的群組

語法：

```
--principal DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:saml-
provider/<SAMLproviderName>:group/<group-name>
```

範例：

```
--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 是 Amazon QuickSight 企業版使用者

語法：

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:<region>:<account-
id>:user/<namespace>/<user-name>
```

> **Note**
>
> 對於 *<namespace>*，您必須指定 `default`。

範例：

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:us-
east-1:111122223333:user/default/bi_user1
```

Principal 是 Amazon QuickSight Enterprise Edition 群組

語法：

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:<region>:<account-
id>:group/<namespace>/<group-name>
```

> **Note**
>
> 對於 *<namespace>*，您必須指定 `default`。

範例：

```
--principal DataLakePrincipalIdentifier=arn:aws:quicksight:us-
east-1:111122223333:group/default/data_scientists
```

委託人是 AWS 帳戶

語法:

```
--principal DataLakePrincipalIdentifier=<account-id>
```

範例：

```
--principal DataLakePrincipalIdentifier=111122223333
```

委託人是組織

語法:

```
--principal DataLakePrincipalIdentifier=arn:aws:organizations::<account-
id>:organization/<organization-id>
```

範例：

```
--principal
 DataLakePrincipalIdentifier=arn:aws:organizations::111122223333:organization/o-
abcdefghijkl
```

Principal 是組織單位

語法:

```
--principal DataLakePrincipalIdentifier=arn:aws:organizations::<account-
id>:ou/<organization-id>/<organizational-unit-id>
```

範例：

```
--principal DataLakePrincipalIdentifier=arn:aws:organizations::111122223333:ou/o-
abcdefghijkl/ou-ab00-cdefghij
```

Principal 是 IAM Identity Center 身分使用者或群組

範例：使用者

```
--principal DataLakePrincipalIdentifier=arn:aws:identitystore:::user/<UserID>
```

範例：群組：

```
--principal DataLakePrincipalIdentifier=arn:aws:identitystore:::group/<GroupID>
```

Principal 是 IAM 群組 - **`IAMAllowedPrincipals`**

Lake Formation 會將資料目錄中所有資料庫和資料表的`Super`許可設定為`IAMAllowedPrincipals`預設呼叫的群組。如果此群組許可存在於資料庫或資料表上，則您帳戶中的所有委託人都可以透過 IAM 委託人政策存取資源 AWS Glue。當您開始使用 Lake Formation 許可來保護先前受到 IAM 政策保護的 Data Catalog 資源時，它可提供回溯相容性 AWS Glue。

當您使用 Lake Formation 管理 Data Catalog 資源的許可時，您必須先撤銷資源的`IAMAllowedPrincipals`許可，或選擇將主體和資源加入混合存取模式， Lake Formation 許可才能運作。

範例：

```
--principal DataLakePrincipalIdentifier=IAM_Allowed_Principals
```

Principal 是 IAM 群組 - **`ALLIAMPrincipals`**

當您授予在 Data Catalog 資源上`ALLIAMPrincipals`分組的許可時，帳戶中的每個主體都可以使用 Lake Formation 許可和 IAM 許可來存取 Data Catalog 資源。

範例：

```
--principal DataLakePrincipalIdentifier=123456789012:IAMPrincipals
```

Lake Formation 許可

本節包含您可以授予委託人的可用 Lake Formation 許可。

ALTER

權限	在此資源上授予	承授者也需要
ALTER	DATABASE	glue:UpdateDatabase
ALTER	TABLE	glue:UpdateTable
ALTER	LF-Tag	lakeformation:UpdateLFTag

具有此許可的主體可以變更 Data Catalog 中資料庫或資料表的中繼資料。對於資料表，您可以變更資料欄結構描述並新增資料欄參數。您無法變更中繼資料資料表指向的基礎資料中的資料欄。

如果正在變更的 屬性是已註冊的 Amazon Simple Storage Service (Amazon S3) 位置，委託人必須擁有新位置的資料位置許可。

Example

下列範例會將 ALTER許可授予 AWS 帳戶 1111-2222-3333 retail中資料庫datalake_user1上的使用者。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "ALTER" --resource '{ "Database": {"Name":"retail"}}'
```

Example

下列範例ALTER會授予inventory資料庫 中 資料表datalake_user1上的使用者 retail。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "ALTER" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

CREATE_DATABASE

權限	在此資源上授予	承授者也需要
CREATE_DATABASE	Data Catalog	glue:CreateDatabase

具有此許可的主體可以在 Data Catalog 中建立中繼資料資料庫或資源連結。委託人也可以在資料庫中建立資料表。

Example

下列範例CREATE_DATABASE授予 AWS 帳戶 1111-2222-3333 datalake_user1中的使用者。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "CREATE_DATABASE" --resource '{ "Catalog": {}}'
```

當委託人在 Data Catalog 中建立資料庫時，不會授予基礎資料的許可。已授予下列其他中繼資料許可（以及將這些許可授予他人的能力）：

- CREATE_TABLE 資料庫中的
- ALTER 資料庫
- DROP 資料庫

建立資料庫時，主體可以選擇性地指定 Amazon S3 位置。根據委託人是否具有資料位置許可，CREATE_DATABASE許可可能不足以在所有情況下建立資料庫。請務必謹記下列三個案例。

建立資料庫使用案例	需要的許可
未指定位置屬性。	CREATE_DATABASE 已足夠。
已指定位置屬性，且位置不是由 Lake Formation 管理（未註冊）。	CREATE_DATABASE 已足夠。
已指定位置屬性，且位置由 Lake Formation 管理（已註冊）。	CREATE_DATABASE 是必要項目，加上指定位置上的資料位置許可。

CREATE_TABLE

權限	在此資源上授予	承授者也需要
CREATE_TABLE	DATABASE	glue:CreateTable

具有此許可的主體可以在指定資料庫中的 Data Catalog 中建立中繼資料表或資源連結。

Example

下列範例授予使用者在 AWS 帳戶 1111-2222-3333 的retail資料庫中建立資料表的datalake_user1許可。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "CREATE_TABLE" --resource '{ "Database": {"Name":"retail"}}'
```

當委託人在 Data Catalog 中建立資料表時，資料表上的所有 Lake Formation 許可都會授予委託人，並能夠將這些許可授予其他人。

跨帳戶授權

如果資料庫擁有者帳戶CREATE_TABLE授予收件人帳戶，且收件人帳戶中的使用者成功在擁有者帳戶的資料庫中建立資料表，則適用下列規則：

- 收件人帳戶中的使用者和資料湖管理員擁有資料表上的所有 Lake Formation 許可。他們可以將資料表的許可授予其帳戶中的其他主體。他們無法將許可授予擁有者帳戶或任何其他帳戶中的主體。
- 擁有者帳戶中的資料湖管理員可以將資料表的許可授予其帳戶中的其他主體。

資料位置許可

當您嘗試建立指向 Amazon S3 位置的資料表時，取決於您是否具有資料位置許可，CREATE_TABLE許可可能不足以建立資料表。請務必謹記下列三個案例。

建立資料表使用案例	需要的許可
指定的位置不是由 Lake Formation 管理 （未註冊）。	CREATE_TABLE 已足夠。

建立資料表使用案例	需要的許可
指定的位置由 Lake Formation （已註冊） 管理，且包含的資料庫沒有位置屬性，或具有不是資料表位置 Amazon S3 字首的位置屬性。	CREATE_TABLE 是必要項目，加上指定位置上的資料位置許可。
指定的位置由 Lake Formation （已註冊） 管理，且包含 的資料庫具有指向已註冊位置的位置屬性，並且是資料表位置的 Amazon S3 字首。	CREATE_TABLE 已足夠。

DATA_LOCATION_ACCESS

權限	在此資源上授予	承授者也需要
DATA_LOCATION_ACCESS	Amazon S3 位置	（位置上的 Amazon S3 許可，必須由用來註冊位置的角色指定。）

這是唯一的資料位置許可。具有此許可的主體可以建立指向指定 Amazon S3 位置的中繼資料資料庫或資料表。位置必須註冊。在位置具有資料位置許可的委託人在子位置也有位置許可。

Example

下列範例`s3://products/retail`會將 上的資料位置許可授予帳戶 1111-2222-3333 `datalake_user1`中的 AWS 使用者。

```
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 不需要查詢或更新基礎資料。此許可僅適用於建立 Data Catalog 資源。

如需資料位置許可的詳細資訊，請參閱 Underlying data access control。

DELETE

權限	在此資源上授予	承授者也需要
DELETE	TABLE	（如果位置已註冊，則不需要額外的 IAM 許可。）

具有此許可的主體可以在資料表指定的 Amazon S3 位置插入、更新和讀取基礎資料。委託人也可以在 Lake Formation 主控台上檢視資料表，並使用 AWS Glue API 擷取資料表的相關資訊。

Example

下列範例會將 DELETE許可授予retail AWS 帳戶 1111-2222-3333 inventory 中資料庫資料表datalake_user1上的使用者。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DELETE" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

此許可僅適用於 Amazon S3 中的資料，不適用於 Amazon Relational Database Service (Amazon RDS) 等其他資料存放區中的資料。

DESCRIBE

權限	在此資源上授予	承授者也需要
DESCRIBE	資料表資源連結	glue:GetTable
	資料庫資源連結	glue:GetDatabase
DESCRIBE	DATABASE	glue:GetDatabase
DESCRIBE	TABLE	glue:GetTable
DESCRIBE	LF-Tag	glue:GetTable
		glue:GetDatabase

權限	在此資源上授予	承授者也需要

		`lakeformation:GetResourceLFTags`
		
		`lakeformation:ListLFTags`
		
		`lakeformation:GetLFTag`
		
		`lakeformation:SearchTablesByLFTags`
		
		`lakeformation:SearchDatabasesByLFTags`

具有此許可的主體可以檢視指定的資料庫、資料表或資源連結。不會隱含授予其他 Data Catalog 許可，也不會隱含授予資料存取許可。資料庫和資料表會出現在整合服務的查詢編輯器中，但除非授予其他 Lake Formation 許可 （例如 SELECT)，否則無法對其進行任何查詢。

例如，DESCRIBE在資料庫中具有 的使用者可以看到資料庫和所有資料庫中繼資料 （描述、位置等）。不過，使用者無法得知資料庫包含哪些資料表，也無法捨棄、變更或建立資料庫中的資料表。同樣地，在資料表DESCRIBE上具有 的使用者可以看到資料表和資料表中繼資料 （描述、結構描述、位置等），但無法捨棄、變更或對資料表執行查詢。

以下是 的一些其他規則DESCRIBE：

- 如果使用者對資料庫、資料表或資源連結具有其他 Lake Formation 許可，DESCRIBE則會隱含授予 。
- 如果使用者SELECT只有資料表的欄子集 （部分 SELECT)，則使用者只能看到這些欄。
- 您無法DESCRIBE將 授予在資料表上有部分選取的使用者。相反地，您無法為DESCRIBE授予 的資料表指定資料欄包含或排除清單。

Example

下列範例會將DESCRIBE許可授予帳戶 `retail` AWS 1111-2222-3333 `inventory-link` 中資料庫`datalake_user1`的資料表資源連結上的使用者。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DESCRIBE" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory-link"}}'
```

DROP

權限	在此資源上授予	承授者也需要
DROP	DATABASE	glue:DeleteDatabase
DROP	TABLE	glue:DeleteTable
DROP	LF-Tag	lakeformation:DeleteLFTag
DROP	資料庫資源連結 資料表資源連結	glue:DeleteDatabase glue:DeleteTable

具有此許可的主體可以在 Data Catalog 中捨棄資料庫、資料表或資源連結。您無法將資料庫上的 DROP 授予外部帳戶或組織。

> **Warning**
>
> 捨棄資料庫會捨棄資料庫中的所有資料表。

Example

下列範例會將 DROP許可授予 AWS 帳戶 1111-2222-3333 `retail`中資料庫`datalake_user1`的使用者。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "DROP" --resource '{ "Database": {"Name":"retail"}}'
```

Example

下列範例DROP會授予資料庫 中資料表 inventory datalake_user1上的使用者 retail。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DROP" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

Example

下列範例會將資料庫 datalake_user1中資料表資源連結inventory-link上的 DROP授予使用者retail。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "DROP" --resource '{ "Table": {"DatabaseName":"retail", "Name":"inventory-
link"}}'
```

INSERT

權限	在此資源上授予	承授者也需要
INSERT	TABLE	（如果位置已註冊，則不需要額外的 IAM 許可。）

具有此許可的主體可以在資料表指定的 Amazon S3 位置插入、更新和讀取基礎資料。委託人也可以在 Lake Formation 主控台中檢視資料表，並使用 AWS Glue API 擷取資料表的相關資訊。

Example

下列範例會將 INSERT許可授予 AWS 帳戶 1111-2222-3333 inventory 中資料庫datalake_user1資料表retail上的使用者。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "INSERT" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

此許可僅適用於 Amazon S3 中的資料，不適用於 Amazon RDS 等其他資料存放區中的資料。

SELECT

權限	在此資源上授予	承授者也需要
SELECT	• TABLE	（如果位置已註冊，則不需要額外的 IAM 許可。）

具有此許可的主體可以檢視 Data Catalog 中的資料表，也可以在資料表指定的位置查詢 Amazon S3 中的基礎資料。委託人可以在 Lake Formation 主控台中檢視資料表，並使用 AWS Glue API 擷取資料表的相關資訊。如果在授予此許可時套用資料欄篩選，委託人只能檢視包含資料欄的中繼資料，並且只能從包含的資料欄查詢資料。

> **Note**
>
> 整合式分析服務負責在處理查詢時套用資料欄篩選。

Example

下列範例會將 SELECT許可授予 AWS 帳戶 1111-2222-3333 `inventory` 中資料庫`datalake_user1`資料表`retail`上的使用者。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "SELECT" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

此許可僅適用於 Amazon S3 中的資料，不適用於 Amazon RDS 等其他資料存放區中的資料。

您可以使用選用的包含清單或排除清單來篩選（限制對 的存取）特定資料欄。包含清單會指定可存取的資料欄。排除清單會指定無法存取的資料欄。如果沒有包含或排除清單，則可以存取所有資料表資料欄。

的結果只會`glue:GetTable`傳回發起人有權檢視的資料欄。整合服務，例如 Amazon Athena 和 Amazon Redshift 遵守資料欄包含和排除清單。

Example

下列範例`inventory`會使用包含清單SELECT將 授予資料表`datalake_user1`上的使用者。

```
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

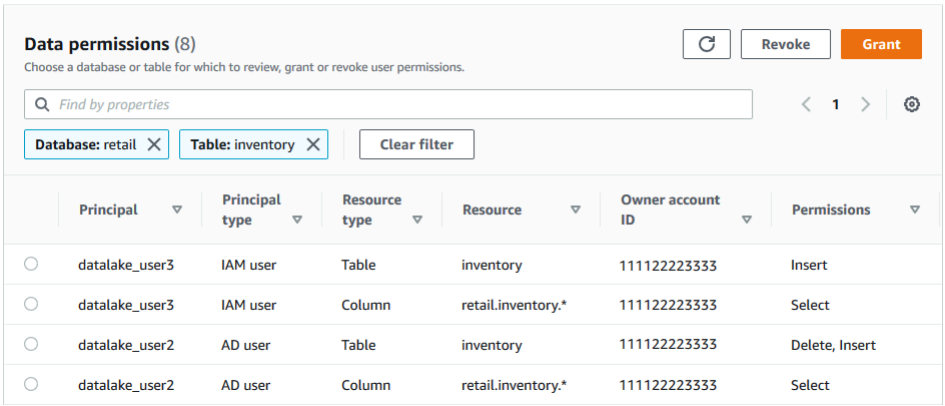
下一個範例會使用排除清單在inventory資料表SELECT上授予 。

```
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"]}}}'
```

下列限制適用於 SELECT許可：

- 授予 時SELECT，如果套用資料欄篩選，則無法包含授予選項。
- 對於屬於分割區索引鍵的資料欄，您無法限制存取控制。
- 在資料表中資料欄子集上具有 SELECT許可的主體，無法獲得該資料表上的 ALTER、DELETE、DROP或 INSERT許可。同樣地，無法在資料表上授予具有 ALTER、DELETE、 DROP或 INSERT許可的主體具有資料欄篩選的SELECT許可。

SELECT 許可一律會以個別資料列的形式出現在 Lake Formation 主控台的資料許可頁面上。下圖顯示 SELECT 已授予inventory資料表中datalake_user3所有資料欄的使用者datalake_user2和 。



Super

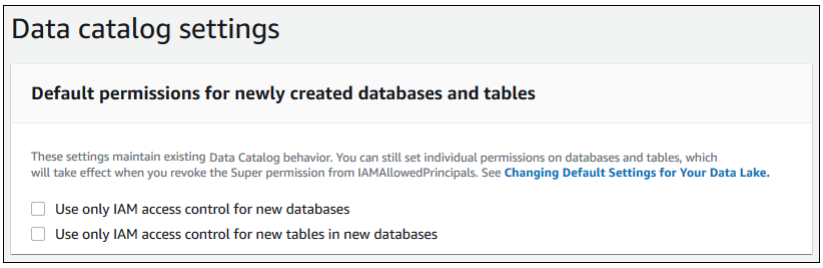
權限	在此資源上授予	承授者也需要
Super	DATABASE	glue:*Database*
Super	TABLE	glue:*Table*, glue:*Partition*

此許可允許主體在資料庫或資料表上執行每個支援的 Lake Formation 操作。您無法將資料庫Super上的 授予外部帳戶。

此許可可以與其他 Lake Formation 許可共存。例如，您可以授予中繼資料資料表上的 SELECT、Super和 INSERT許可。然後，委託人可以在資料表上執行所有支援的操作。當您撤銷 時Super，SELECT和 INSERT許可會保留，委託人只能執行選取和插入操作。

您可以將其授予群組 ，而不是Super授予個別委託人IAMAllowedPrincipals。IAMAllowedPrincipals 群組會自動建立，並包含 IAM 政策允許存取 Data Catalog 資源的所有 IAM 使用者和角色。Super 將 Data Catalog 資源IAMAllowedPrincipals的存取權授予 時，只有 IAM 政策才能有效控制資源的存取。

您可以利用 Lake Formation 主控台的設定頁面上的選項，將IAMAllowedPrincipals新目錄資源Super的許可自動授予 。



- 若要IAMAllowedPrincipals為所有新資料庫授予 Super ，請選取僅對新資料庫使用 IAM 存取控制。
- 若要IAMAllowedPrincipals為新資料庫中的所有新資料表授予 Super ，請選取僅對新資料庫中的新資料表使用 IAM 存取控制。

> **Note**
>
> 此選項會導致核取方塊 在建立資料庫對話方塊中，依預設只針對此資料庫中的新資料表使用 IAM 存取控制。 它什麼都不做。這是建立資料庫對話方塊中的核取方塊，可讓 授予 `Super`給 `IAMAllowedPrincipals`。

這些設定頁面選項預設為啟用。如需詳細資訊，請參閱下列內容：

- the section called “變更資料湖的預設設定”
- the section called “將AWS Glue資料許可升級至 Lake Formation 模型”

SUPER_USER

權限	在此資源上授予	承授者也需要
`Super user`	`Catalog`	`glue:GetCatalog`

您只能將 `Super user`許可授予預設 Data Catalog 內目錄上的特定主體。您無法將預設目錄或其他資源類型的`Super user`許可授予外部帳戶中的主體，例如資料庫和資料表。`Super user` 許可許可允許委託人在授予目錄中的資料庫和資料表上執行每個支援的 Lake Formation 操作。

透過 `Super user`許可，委託人（承授者）能夠對目錄中的資源（目錄、資料庫和資料表）執行下列動作：

- `CREATE_DATABASE`，目錄上的`DESCRIBE`許可。
- `DROP`目錄內所有資料庫的 `ALTER`、、`CREATE_TABLE`、 `DESCRIBE` (有效 `SUPER`) 許可。
- `DROP`目錄內所有資料庫內所有資料表的 `ALTER`、`DESCRIBE`、`SELECT`、`INSERT`、、 `DELETE` (有效 `SUPER`) 許可。
- `All` (有效 `SUPER`) 目錄內目錄的許可。
- 授予（將這些許可授予其他主體的能力）目錄內所有目錄、資料庫和資料表的許可。

透過目錄資源的 `Super user` 許可，承授者不得在目錄上執行或委派 `ALTER`和 `DROP`動作。

ASSOCIATE

權限	在此資源上授予	承授者也需要
ASSOCIATE	LF-Tag	glue:GetDatabase glue:GetTable lakeformation:AddLFTagsToResource" lakeformation:RemoveLFTagsFromResource" lakeformation:GetResourceLFTags lakeformation:ListLFTags lakeformation:GetLFTag lakeformation:SearchTablesByLFTags lakeformation:SearchDatabasesByLFTags

在 LF-Tag 具有此許可的主體可以將 LF-Tag 指派給 Data Catalog 資源。ASSOCIATE 隱含授予 DESCRIBE。

Example

此範例datalake_user1使用金鑰 授予使用者 LF-Tag 上的ASSOCIATE許可module。它授予檢視和指派該索引鍵所有值的許可，如星號 (*) 所示。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --permissions "ASSOCIATE" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

整合 IAM Identity Center

使用 AWS IAM Identity Center，您可以連線至身分提供者 IdPs)，並跨 AWS 分析服務集中管理使用者和群組的存取權。您可以將 Okta、Ping 和 Microsoft Entra ID （先前稱為 Azure Active Directory) 等身分提供者與 IAM Identity Center 整合，讓組織中的使用者使用單一登入體驗來存取資料。IAM Identity Center 也支援連接其他第三方身分提供者。

如需詳細資訊，請參閱 AWS IAM Identity Center 《使用者指南》中的支援的身分提供者。

您可以在 IAM Identity Center 中將 AWS Lake Formation 設定為已啟用的應用程式，而資料湖管理員可以將精細許可授予 AWS Glue Data Catalog 資源上的授權使用者和群組。

您組織的使用者可以使用組織的身分提供者登入任何已啟用 Identity Center 的應用程式，並查詢套用 Lake Formation 許可的資料集。透過此整合，您可以管理對 AWS 服務的存取，而無需建立多個 IAM 角色。

> **Note**
>
> 信任的身分傳播允許使用者的現有使用者和群組成員資格跨 AWS 分析服務存取資料。透過信任的身分傳播，使用者可以登入應用程式，而應用程式可以在請求中傳遞使用者的身分，以存取 AWS 服務中的資料。您不需要執行任何服務特定的身分提供者組態或 IAM 角色設定。使用者無法使用信任 AWS Management Console 的身分傳播登入 。如需詳細資訊，請參閱 AWS IAM Identity Center 《使用者指南》中的跨應用程式的受信任身分傳播。

如需限制的詳細資訊，請參閱IAM Identity Center 整合限制。

主題

- IAM Identity Center 與 Lake Formation 整合的先決條件
- 將 Lake Formation 與 IAM Identity Center 連線
- 更新 IAM Identity Center 整合
- 刪除 Lake Formation 與 IAM Identity Center 的連線
- 將許可授予使用者和群組
- 在 CloudTrail 日誌中包含 IAM Identity Center 使用者內容

IAM Identity Center 與 Lake Formation 整合的先決條件

以下是整合 IAM Identity Center 與 Lake Formation 的先決條件。

1. 啟用 IAM Identity Center – 啟用 IAM Identity Center 是支援身分驗證和身分傳播的先決條件。
2. 選擇您的身分來源 - 啟用 IAM Identity Center 之後，您必須有身分提供者才能管理使用者和群組。您可以使用內建的 Identity Center 目錄做為身分來源，或使用外部 IdP，例如 Microsoft Entra ID 或 Okta。

   如需詳細資訊，請參閱 AWS IAM Identity Center 《使用者指南》中的管理身分來源和連線至外部身分提供者。
3. 建立 IAM 角色 – 建立 IAM Identity Center 連線的角色需要許可，才能在 Lake Formation 和 IAM Identity Center 中建立和修改應用程式組態，如下列內嵌政策所示。

   您需要根據 IAM 最佳實務新增許可。下列程序會詳細說明特定權限。如需詳細資訊，請參閱 IAM Identity Center 入門。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:CreateLakeFormationIdentityCenterConfiguration",
                "sso:CreateApplication",
                "sso:PutApplicationAssignmentConfiguration",
                "sso:PutApplicationAuthenticationMethod",
                "sso:PutApplicationGrant",
                "sso:PutApplicationAccessScope",
            ],
            "Resource": [
                "*"
            ]
        }
    ]
}
```

如果您要與外部 AWS 帳戶 或組織共用 Data Catalog 資源，您必須具有建立資源共用的 AWS Resource Access Manager (AWS RAM) 許可。如需共用資源所需許可的詳細資訊，請參閱跨帳戶資料共用先決條件。

下列內嵌政策包含檢視、更新和刪除與 IAM Identity Center 整合之 Lake Formation 屬性所需的特定許可。

- 使用下列內嵌政策，允許 IAM 角色檢視與 IAM Identity Center 的 Lake Formation 整合。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:DescribeLakeFormationIdentityCenterConfiguration",
                "sso:DescribeApplication"
            ],
            "Resource": [
                "*"
            ]
        }
    ]
}
```

- 使用下列內嵌政策，允許 IAM 角色更新與 IAM Identity Center 的 Lake Formation 整合。此政策也包含與外部帳戶共用資源所需的選用許可。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:UpdateLakeFormationIdentityCenterConfiguration",
                "lakeformation:DescribeLakeFormationIdentityCenterConfiguration",
                "sso:DescribeApplication",
                "sso:UpdateApplication",
            ],
            "Resource": [
                "*"
```

```
            ]
        }
    ]
}
```

- 使用下列內嵌政策，允許 IAM 角色刪除與 IAM Identity Center 的 Lake Formation 整合。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:DeleteLakeFormationIdentityCenterConfiguration",
                "sso:DeleteApplication",
            ],
            "Resource": [
                "*"
            ]
        }
    ]
}
```

- 如需為 IAM Identity Center 使用者和群組授予或撤銷資料湖許可所需的 IAM 許可，請參閱 授予或撤銷 Lake Formation 許可所需的 IAM 許可。

許可描述

- `lakeformation:CreateLakeFormationIdentityCenterConfiguration` – 建立 Lake Formation IdC 組態。
- `lakeformation:DescribeLakeFormationIdentityCenterConfiguration` – 描述現有的 IdC 組態。
- `lakeformation:DeleteLakeFormationIdentityCenterConfiguration` – 提供刪除現有 Lake Formation IdC 組態的能力。
- `lakeformation:UpdateLakeFormationIdentityCenterConfiguration` – 用來變更現有的 Lake Formation 組態。
- `sso:CreateApplication`— 用於建立 IAM Identity Center 應用程式。
- `sso:DeleteApplication`— 用於刪除 IAM Identity Center 應用程式。

- `sso:UpdateApplication`— 用於更新 IAM Identity Center 應用程式。
- `sso:PutApplicationGrant`— 用於變更受信任的字符發行者資訊。
- `sso:PutApplicationAuthenticationMethod` – 授予 Lake Formation 身分驗證存取權。
- `sso:GetApplicationGrant`— 用於列出受信任的字符發行者資訊。
- `sso:DeleteApplicationGrant` – 刪除信任權杖發行者資訊。
- `sso:PutApplicationAccessScope` – 新增或更新應用程式 IAM Identity Center 存取範圍的授權目標清單。
- `sso:PutApplicationAssignmentConfiguration` – 用來設定使用者如何存取應用程式。

將 Lake Formation 與 IAM Identity Center 連線

您必須先完成下列步驟，才能使用 IAM Identity Center 管理身分，以使用 Lake Formation 授予對 Data Catalog 資源的存取權。您可以使用 Lake Formation 主控台或 建立 IAM Identity Center 整合 AWS CLI。

AWS Management Console

將 Lake Formation 與 IAM Identity Center 連線

1. 登入 AWS Management Console，然後開啟 Lake Formation 主控台，網址為 https：//https://console.aws.amazon.com/lakeformation/。
2. 在左側導覽窗格中，選取 IAM Identity Center 整合。

Create IAM Identity Center Integration

Enable IAM Identify Center and then create Lake Formation - IAM Identity Center integration to manage identities from IAM Identity Center (external IdPs like Azure AD or Okta Universal Directory). Learn more

How it works

Enable IAM Identity Center
Enable IAM Identity Center for your account or organization and select an identity provider.

Create Lake Formation integration
Integrate Lake Formation with IAM Identity Center to permit Lake Formation to access users from your selected identity provider.

Grant permissions
Grant permissions to users on Data Catalog databases and tables using fine-grained Lake Formation permissions.

Connect Lake Formation to IAM Identity Center

Connect to organization instance of IAM Identity Center

Manage access to Lake Formation by assigning users and groups from the Identity Center directory for your organization. Learn more

Recommended

Connect to account instance of IAM Identity Center

Manage access to Lake Formation by assigning existing or creating dedicated users and groups from your Identity Center directory. Learn more

instance of IAM Identity Center
Manage access to Lake Formation by assigning users and groups from your Identity Center directory.
arn:aws:sso:::instance/ssoins-6987513bf5410c2f

Add AWS account and organization IDs

Add AWS accounts and organizations whose users need access to Lake Formation managed resources.

AWS Accounts and AWS organizations
Enter one or more AWS account IDs and AWS organization IDs. Press Enter after each ID.
Choose AWS account, AWS organization ID

Lake Formation application integration - *optional*

Add application ARNs that can access S3 data locations registered with Lake Formation on behalf of the user.

After this step, you can't edit the connection. You can edit AWS accounts, organizations, and applications. If you want to modify the connection, delete it and create a new connection.

3. （選用）輸入一或多個 AWS 帳戶 IDs、組織 IDs和/或組織單位 IDs，以允許外部帳戶存取 Data Catalog 資源。當 IAM Identity Center 使用者或群組嘗試存取 Lake Formation 受管 Data Catalog 資源時，Lake Formation 會擔任 IAM 角色來授權中繼資料存取。如果 IAM 角色屬於沒有 AWS Glue 資源政策和 AWS RAM 資源共用的外部帳戶，IAM Identity Center 使用者和群組將無法存取資源，即使他們具有 Lake Formation 許可。

   Lake Formation 使用 AWS Resource Access Manager (AWS RAM) 服務與外部帳戶和組織共用資源。 AWS RAM 向承授者帳戶傳送邀請，以接受或拒絕資源共用。

   如需詳細資訊，請參閱從 接受資源共享邀請 AWS RAM。

   > **Note**
   >
   > Lake Formation 允許來自外部帳戶的 IAM 角色代表 IAM Identity Center 使用者和群組擔任電信業者角色，以存取 Data Catalog 資源，但只能在擁有帳戶中的 Data Catalog 資源上授予許可。如果您嘗試將許可授予外部帳戶中 Data Catalog 資源上的 IAM Identity Center 使用者和群組，Lake Formation 會擲出下列錯誤 -「委託人不支援跨帳戶授予。」

4. （選用）在建立 Lake Formation 整合畫面上，指定第三方應用程式的 ARNs，這些應用程式可存取向 Lake Formation 註冊的 Amazon S3 位置中的資料。Lake Formation 會根據有效許可，以 AWS STS 字符的形式將範圍縮小的臨時憑證提供給已註冊的 Amazon S3 位置，以便授權的應用程式可以代表使用者存取資料。
5. 選取提交。

   Lake Formation 管理員完成步驟並建立整合後，IAM Identity Center 屬性會出現在 Lake Formation 主控台中。完成這些任務可讓 Lake Formation 成為啟用 IAM Identity Center 的應用程式。主控台內的屬性包括整合狀態。整合狀態會顯示完成`Success`的時間。此狀態指出 IAM Identity Center 組態是否已完成。

AWS CLI

- 下列範例示範如何建立 Lake Formation 與 IAM Identity Center 的整合。您也可以指定應用程式的 `Status`(`ENABLED`、`DISABLED`)。

```
aws lakeformation create-lake-formation-identity-center-configuration \
    --catalog-id <123456789012> \
    --instance-arn <arn:aws:sso:::instance/ssoins-112111f12ca1122p> \
```

```
    --share-recipients '[{"DataLakePrincipalIdentifier": "<123456789012>"},
                         {"DataLakePrincipalIdentifier": "<555555555555>"}]' \
    --external-filtering '{"AuthorizedTargets": ["<app arn1>", "<app arn2>"],
 "Status": "ENABLED"}'
```

- 下列範例顯示如何檢視 Lake Formation 與 IAM Identity Center 的整合。

```
aws lakeformation describe-lake-formation-identity-center-configuration
 --catalog-id <123456789012>
```

更新 IAM Identity Center 整合

建立連線後，您可以新增第三方應用程式，讓 IAM Identity Center 整合與 Lake Formation 整合，並代表使用者存取 Amazon S3 資料。您也可以從 IAM Identity Center 整合中移除現有的應用程式。您可以使用 Lake Formation 主控台，以及 UpdateLakeFormationIdentityCenterConfiguration 操作 AWS CLI 來新增或移除應用程式。

> **Note**
>
> 建立 IAM Identity Center 整合之後，您無法更新執行個體 ARN。

AWS Management Console

更新與 Lake Formation 的現有 IAM Identity Center 連線

1. 登入 AWS Management Console，然後開啟 Lake Formation 主控台，網址為 https：//https://console.aws.amazon.com/lakeformation/。
2. 在左側導覽窗格中，選取 IAM Identity Center 整合。
3. 在 IAM Identity Center 整合頁面上選取新增。
4. 輸入一或多個 AWS 帳戶 IDs、組織 IDs和/或組織單位 IDs，以允許外部帳戶存取 Data Catalog 資源。
5. 在新增應用程式畫面上，輸入您要與 Lake Formation 整合之第三方應用程式的應用程式 IDs。
6. 選取新增。

AWS CLI

您可以執行下列 AWS CLI 命令，為 IAM Identity Center 整合新增或移除第三方應用程式。當您將外部篩選狀態設定為 時ENABLED，它可讓 IAM Identity Center 為第三方應用程式提供身分管理，以存取 Lake Formation 管理的資料。您也可以設定應用程式狀態來啟用或停用 IAM Identity Center 整合。

```
aws lakeformation update-lake-formation-identity-center-configuration \
 --external-filtering '{"AuthorizedTargets": ["<app arn1>", "<app arn2>"], "Status":
 "ENABLED"}'\
 --share-recipients '[{"DataLakePrincipalIdentifier": "<444455556666>"}
                     {"DataLakePrincipalIdentifier": "<777788889999>"}]' \
 --application-status ENABLED
```

刪除 Lake Formation 與 IAM Identity Center 的連線

如果您想要刪除現有的 IAM Identity Center 整合，您可以使用 Lake Formation 主控台 AWS CLI或 DeleteLakeFormationIdentityCenterConfiguration 操作來執行。

AWS Management Console

刪除與 Lake Formation 的現有 IAM Identity Center 連線

1. 登入 AWS Management Console，然後開啟 Lake Formation 主控台，網址為 https：//https://console.aws.amazon.com/lakeformation/。
2. 在左側導覽窗格中，選取 IAM Identity Center 整合。
3. 在 IAM Identity Center 整合頁面上選取刪除。
4. 在確認整合畫面上，確認動作，然後選取刪除。

AWS CLI

您可以執行下列 AWS CLI 命令來刪除 IAM Identity Center 整合。

```
aws lakeformation delete-lake-formation-identity-center-configuration \
    --catalog-id <123456789012>
```

將許可授予使用者和群組

您的資料湖管理員可以將許可授予 Data Catalog 資源（資料庫、資料表和檢視）上的 IAM Identity Center 使用者和群組，以便輕鬆存取資料。若要授予或撤銷資料湖許可，授予者需要下列 IAM Identity Center 動作的許可。

- DescribeUser
- DescribeGroup
- DescribeInstance

您可以使用 Lake Formation 主控台、API 或 來授予許可 AWS CLI。

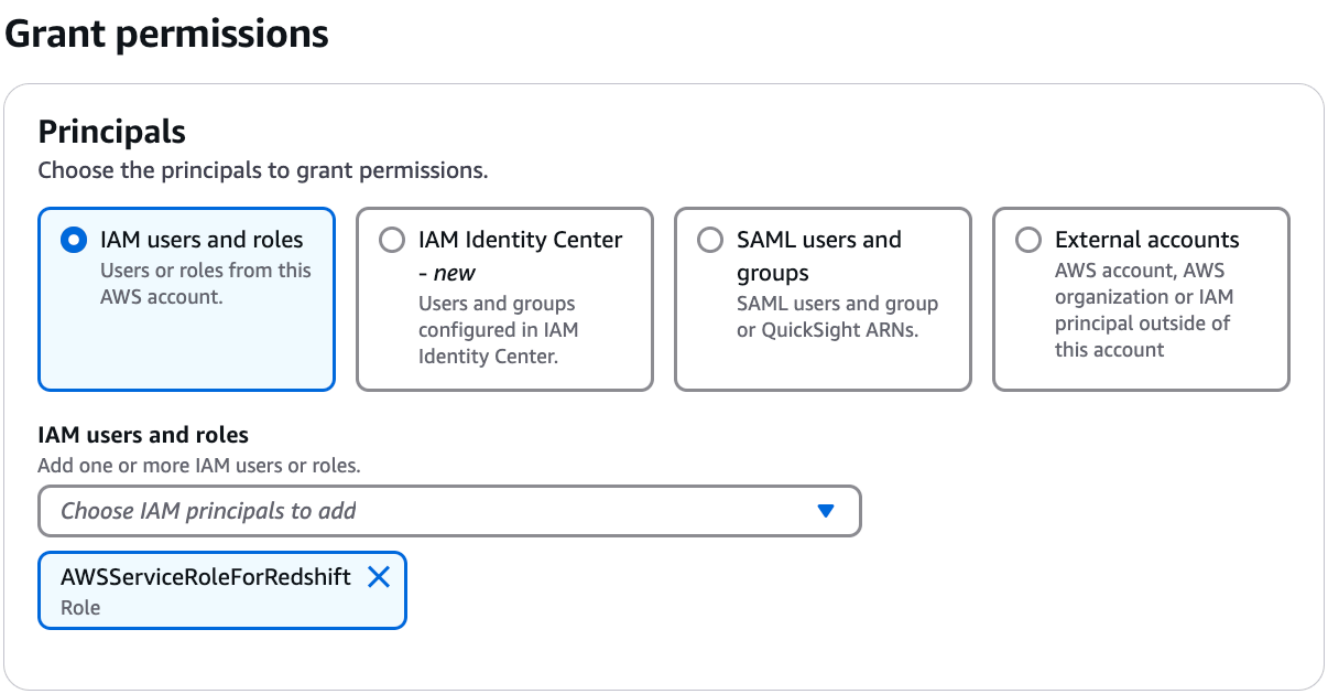
如需授予許可的詳細資訊，請參閱 the section called “授予資料許可”。

> **Note**
> 您只能授予帳戶中資源的許可。若要將許可串聯至與您共用資源上的使用者和群組，您必須使用 AWS RAM 資源共用。

AWS Management Console

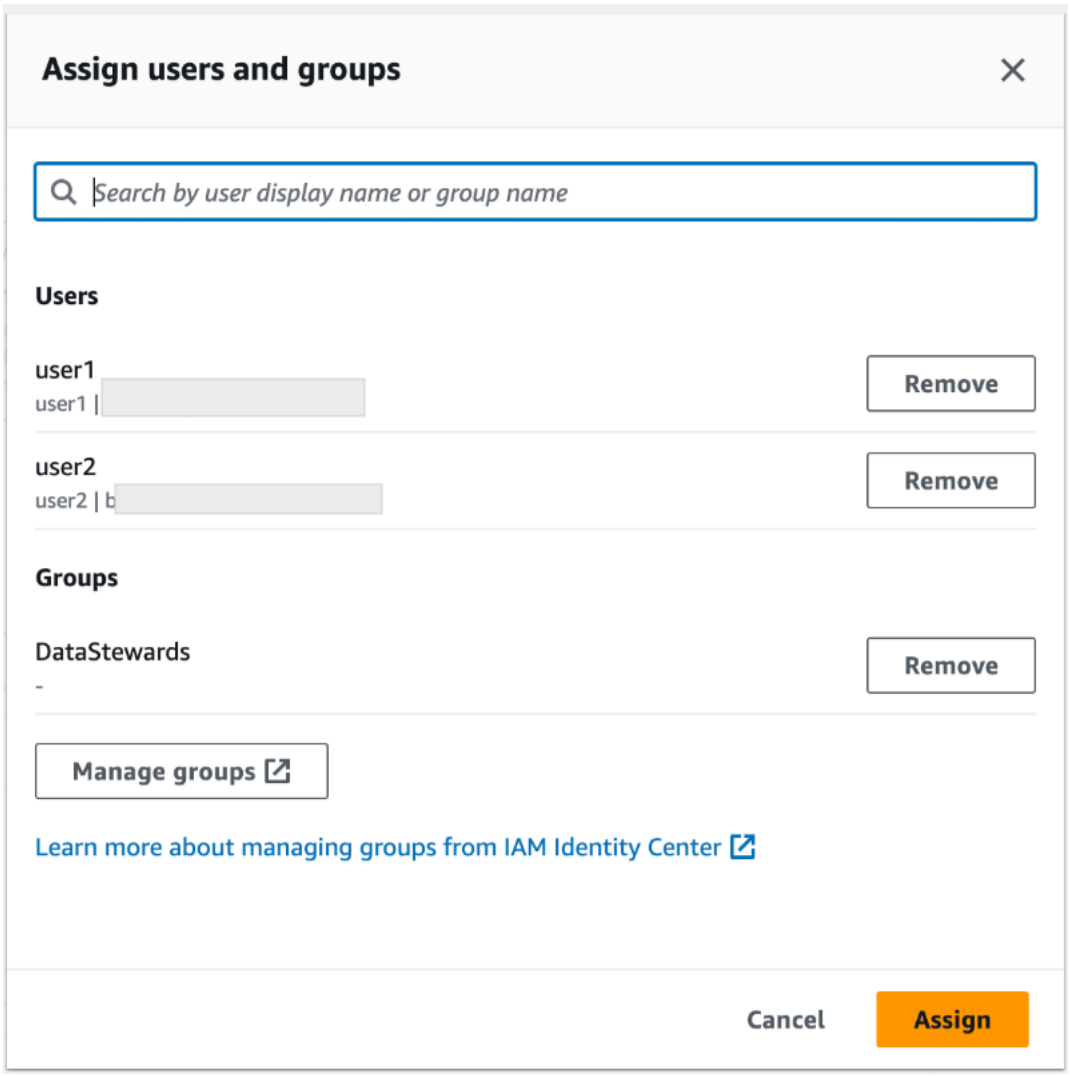
將許可授予使用者和群組

1. 登入 AWS Management Console，然後開啟 Lake Formation 主控台，網址為 https：//https://console.aws.amazon.com/lakeformation/。
2. 在 Lake Formation 主控台的許可下選取資料湖許可。
3. 選取授予。
4. 在授予資料湖許可頁面上，選擇 IAM Identity Center 使用者和群組。
5. 選取新增以選擇要授予許可的使用者和群組。



6. 在指派使用者和群組畫面上，選擇要授予許可的使用者和/或群組。

   選取指派。



7. 接著，選擇授予許可的方法。

   如需使用具名資源方法授予許可的說明，請參閱 使用具名資源方法授予資料許可。

   如需使用 LF 標籤授予許可的說明，請參閱 使用 LF-TBAC 方法授予資料湖許可。
8. 選擇您要授予許可的 Data Catalog 資源。
9. 選擇要授予的資料目錄許可。
10. 選取授予。

AWS CLI

下列範例顯示如何授予資料表的 IAM Identity Center 使用者SELECT許可。

```
aws lakeformation grant-permissions \
```

```
--principal DataLakePrincipalIdentifier=arn:aws:identitystore:::user/<UserId> \
--permissions "SELECT" \
--resource '{ "Table": { "DatabaseName": "retail", "TableWildcard": {} } }'
```

若要UserId從 IAM Identity Center 擷取，請參閱《IAM Identity Center API 參考》中的 GetUserId 操作。

在 CloudTrail 日誌中包含 IAM Identity Center 使用者內容

Lake Formation 使用憑證販賣功能來提供暫時存取 Amazon S3 資料。根據預設，當 IAM Identity Center 使用者提交查詢至整合式分析服務時，CloudTrail 日誌只會包含服務擔任的 IAM 角色，以提供短期存取。如果您使用使用者定義的角色向 Lake Formation 註冊 Amazon S3 資料位置，您可以選擇在 CloudTrail 事件中包含 IAM Identity Center 使用者的內容，然後追蹤存取 資源的使用者。

> **Important**
>
> 若要在 CloudTrail 中包含物件層級的 Amazon S3 API 請求，您需要為 Amazon S3 儲存貯體和物件啟用 CloudTrail 事件記錄。如需更多內建，請參閱《Amazon S3 使用者指南》中的啟用 Amazon S3 儲存貯體和物件的 CloudTrail 事件記錄。 Amazon S3

在向使用者定義角色註冊的資料湖位置上啟用憑證販賣稽核

1. 登入 Lake Formation 主控台，網址為 https：//https://console.aws.amazon.com/lakeformation/。
2. 在左側導覽中，展開管理，然後選擇資料目錄設定。
3. 在增強型稽核下，選擇傳播提供的內容。
4. 選擇 Save (儲存)。

您也可以在 PutDataLakeSettings 操作中設定 Parameters 屬性，以啟用增強型稽核選項。根據預設，SET_CONTEXT"參數值設定為「true」。

```
{
    "DataLakeSettings": {
        "Parameters": {"SET_CONTEXT": "true"},
    }
```

```
}
```

以下是 CloudTrail 事件的摘錄，其中包含增強型稽核選項。此日誌同時包含 IAM Identity Center 使用者的工作階段內容，以及 Lake Formation 擔任以存取 Amazon S3 資料位置的使用者定義 IAM 角色。請參閱下列摘錄中的 onBehalfOf 參數。

```
{
         "eventVersion":"1.09",
         "userIdentity":{
            "type":"AssumedRole",
            "principalId":"AROAW7F7MOX4OYE6FLIFN:access-grants-
e653760c-4e8b-44fd-94d9-309e035b75ab",
            "arn":"arn:aws:sts::123456789012:assumed-role/accessGrantsTestRole/access-
grants-e653760c-4e8b-44fd-94d9-309e035b75ab",
            "accountId":"123456789012",
            "accessKeyId":"ASIAW7F7MOX4CQLD4JIZN",
            "sessionContext":{
               "sessionIssuer":{
                  "type":"Role",
                  "principalId":"AROAW7F7MOX4OYE6FLIFN",
                  "arn":"arn:aws:iam::123456789012:role/accessGrantsTestRole",
                  "accountId":"123456789012",
                  "userName":"accessGrantsTestRole"
               },
               "attributes":{
                  "creationDate":"2023-08-09T17:24:02Z",
                  "mfaAuthenticated":"false"
               }
            },
            "onBehalfOf":{
                "userId": "<identityStoreUserId>",
                "identityStoreArn": "arn:aws:identitystore::<restOfIdentityStoreArn>"
            }
         },
         "eventTime":"2023-08-09T17:25:43Z",
         "eventSource":"s3.amazonaws.com",
         "eventName":"GetObject",
    ....
```

將 Amazon S3 位置新增至您的資料湖

若要將資料位置新增為資料湖中的儲存體，您可以向 註冊位置 （資料湖位置) AWS Lake Formation。然後，您可以使用 Lake Formation 許可，對指向此位置的 AWS Glue Data Catalog 物件和位置中基礎資料的精細存取控制。

Lake Formation 也允許 以混合存取模式註冊資料位置，並提供您在 Data Catalog 中選擇性地為資料庫和資料表啟用 Lake Formation 許可的彈性。使用混合存取模式時，您會有一個增量路徑，可讓您為一組特定使用者設定 Lake Formation 許可，而不會中斷其他現有使用者或工作負載的許可政策。

如需設定混合存取模式的詳細資訊，請參閱 混合存取模式

當您註冊位置時，該 Amazon S3 路徑和該路徑下的所有資料夾都會註冊。

例如，假設您有如下所示的 Amazon S3 路徑組織：

```
/mybucket/accounting/sales/
```

如果您註冊 `S3://mybucket/accounting`，`sales`資料夾也會在 Lake Formation 管理下註冊。

如需註冊位置的詳細資訊，請參閱 Underlying data access control。

> **Note**
>
> Lake Formation 許可建議用於結構化資料 （在具有資料列和資料欄的資料表中排列）。如果您的資料包含以物件為基礎的非結構化資料，請考慮使用 Amazon S3 存取授權來管理資料存取。

主題

- 用於註冊位置的角色需求
- 註冊 Amazon S3 位置
- 註冊加密的 Amazon S3 位置
- 在另一個 AWS 帳戶中註冊 Amazon S3 位置
- 跨 AWS 帳戶註冊加密的 Amazon S3 位置
- 取消註冊 Amazon S3 位置

用於註冊位置的角色需求

註冊 Amazon Simple Storage Service (Amazon S3) 位置時，您必須指定 AWS Identity and Access Management (IAM) 角色。 會在存取該位置中的資料時 AWS Lake Formation 擔任該角色。

您可以使用下列其中一個角色類型來註冊位置：

- Lake Formation 服務連結角色。此角色會授予位置所需的許可。使用此角色是註冊位置的最簡單方法。如需詳細資訊，請參閱使用 Lake Formation 的服務連結角色及服務連結角色限制。
- 使用者定義的角色。當您需要授予比服務連結角色更多的許可時，請使用使用者定義的角色。

  在下列情況下，您必須使用使用者定義的角色：

  - 在另一個帳戶中註冊位置時。

    如需詳細資訊，請參閱the section called “在另一個 AWS 帳戶中註冊 Amazon S3 位置”及the section called “跨 AWS 帳戶註冊加密的 Amazon S3 位置”。
  - 如果您使用 AWS 受管 CMK (`aws/s3`) 來加密 Amazon S3 位置。

    如需詳細資訊，請參閱註冊加密的 Amazon S3 位置。
  - 如果您計劃使用 Amazon EMR 存取位置。

    如果您已使用服務連結角色註冊位置，並想要開始使用 Amazon EMR 存取該位置，則必須取消註冊該位置，並使用使用者定義的角色重新註冊該位置。如需詳細資訊，請參閱the section called “取消註冊 Amazon S3 位置”。

使用 Lake Formation 的服務連結角色

AWS Lake Formation 使用 AWS Identity and Access Management (IAM) 服務連結角色。服務連結角色是直接連結至 Lake Formation 的唯一 IAM 角色類型。服務連結角色由 Lake Formation 預先定義，並包含該服務代表您呼叫其他 AWS 服務所需的所有許可。

服務連結角色可讓您更輕鬆地設定 Lake Formation，因為您不必建立角色並手動新增必要的許可。Lake Formation 定義其服務連結角色的許可，除非另有定義，否則只有 Lake Formation 可以擔任其角色。定義的許可包括信任政策和許可政策，並且該許可政策不能連接到任何其他 IAM 實體。

此服務連結角色信任下列服務擔任該角色：

- `lakeformation.amazonaws.com`

當您使用帳戶 A 中的服務連結角色來註冊帳戶 B 擁有的 Amazon S3 位置時，帳戶 B 中的 Amazon S3 儲存貯體政策 （以資源為基礎的政策） 必須授予帳戶 A 中服務連結角色的存取許可。

如需使用服務連結角色註冊資料位置的詳細資訊，請參閱 服務連結角色限制。

> **Note**
>
> 服務控制政策 SCPs) 不會影響服務連結角色。
> 如需詳細資訊，請參閱《 AWS Organizations 使用者指南》中的服務控制政策 (SCPs)。

Lake Formation 的服務連結角色許可

Lake Formation 使用名為 的服務連結角色`AWSServiceRoleForLakeFormationDataAccess`。此角色提供一組 Amazon Simple Storage Service (Amazon S3) 許可，可讓 Lake Formation 整合服務 （例如 Amazon Athena) 存取已註冊的位置。註冊資料湖位置時，您必須提供在該位置具有必要 Amazon S3 讀取/寫入許可的角色。您可以使用此服務連結角色，而不是建立具有所需 Amazon S3 許可的角色。

第一次將服務連結角色命名為註冊路徑的角色時，系統會代表您建立服務連結角色和新的 IAM 政策。Lake Formation 會將路徑新增至內嵌政策，並將其連接至服務連結角色。當您使用服務連結角色註冊後續路徑時，Lake Formation 會將路徑新增至現有政策。

以資料湖管理員身分登入時，請註冊資料湖位置。然後，在 IAM 主控台中搜尋角色`AWSServiceRoleForLakeFormationDataAccess`並檢視其連接的政策。

例如，在您註冊位置 之後`s3://my-kinesis-test/logs`，Lake Formation 會建立下列內嵌政策並將其連接至 `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"
            ]
        }
    ]
}
```

為 Lake Formation 建立服務連結角色

您不需要手動建立一個服務連結角色。當您在 AWS CLI、 或 AWS API 中向 Lake Formation 註冊 Amazon S3 AWS Management Console位置時，Lake Formation 會為您建立服務連結角色。

> **Important**
>
> 此服務連結角色可以顯示在您的帳戶，如果您於其他服務中完成一項動作時，可以使用支援此角色的功能。若要進一步了解，請參閱我的 IAM 帳戶中出現的新角色。

若您刪除此服務連結角色，之後需要再次建立，您可以在帳戶中使用相同程序重新建立角色。當您向 Lake Formation 註冊 Amazon S3 位置時，Lake Formation 會再次為您建立服務連結角色。

您也可以使用 IAM 主控台建立具有 Lake Formation 使用案例的服務連結角色。在 AWS CLI 或 AWS API 中，使用服務名稱建立`lakeformation.amazonaws.com`服務連結角色。如需詳細資訊，請參閱《IAM 使用者指南》中的「建立服務連結角色」。如果您刪除此服務連結角色，您可以使用此相同的程序以再次建立該角色。

編輯 Lake Formation 的服務連結角色

Lake Formation 不允許您編輯`AWSServiceRoleForLakeFormationDataAccess`服務連結角色。因為有各種實體可能會參考服務連結角色，所以您無法在建立角色之後變更角色名稱。然而，您可使用 IAM 來編輯角色描述。如需更多資訊，請參閱 IAM 使用者指南中的編輯服務連結角色。

刪除 Lake Formation 的服務連結角色

若您不再使用需要服務連結角色的功能或服務，我們建議您刪除該角色。如此一來，您就沒有未主動監控或維護的未使用實體。然而，在手動刪除服務連結角色之前，您必須先清除資源。

> **Note**
>
> 如果 Lake Formation 服務在您嘗試刪除資源時使用角色，則刪除可能會失敗。若此情況發生，請等待數分鐘後並再次嘗試操作。

刪除 Lake Formation 使用的 Lake Formation 資源

- 如果您已使用服務連結角色向 Lake Formation 註冊 Amazon S3 位置，則在刪除服務連結角色之前，您需要取消註冊該位置，並使用自訂角色重新註冊該位置。

使用 IAM 手動刪除服務連結角色

使用 IAM 主控台 AWS CLI、 或 AWS API 來刪除`AWSServiceRoleForLakeFormationDataAccess`服務連結角色。如需詳細資訊，請參閱《IAM 使用者指南》中的刪除服務連結角色。

以下是使用者定義角色的要求：

- 建立新角色時，在 IAM 主控台的建立角色頁面上，選擇AWS 服務，然後在選擇使用案例下，選擇 Lake Formation。

  如果您使用不同的路徑建立角色，請確定角色與 具有信任關係`lakeformation.amazonaws.com`。如需詳細資訊，請參閱修改角色信任政策（主控台）。
- 角色必須具有內嵌政策，授予 Amazon S3 對該位置的讀取/寫入許可。以下是典型的政策。

```
{
    "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"
            ]
        }
    ]
}
```

- 將下列信任政策新增至 IAM 角色，以允許 Lake Formation 服務擔任角色，並將臨時詳細資訊提供給整合的分析引擎。

  若要在 CloudTrail 日誌中包含 IAM Identity Center 使用者內容，信任政策必須具有 `sts:SetContext`動作的 許可。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "DataCatalogViewDefinerAssumeRole1",
            "Effect": "Allow",
            "Principal": {
               "Service": [
                    "lakeformation.amazonaws.com"
                ]
            },
            "Action": [
                "sts:AssumeRole",
                "sts:SetContext"
            ]
        }
    ]
}
```

- 註冊位置的資料湖管理員必須具有角色的 `iam:PassRole` 許可。

以下是授予此許可的內嵌政策。以有效的 AWS 帳號取代 *<account-id>*，並以角色名稱取代 *<role-name>*。

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

- 若要允許 Lake Formation 在 CloudWatch Logs 中新增日誌並發佈指標，請新增下列內嵌政策。

> **Note**
>
> 寫入 CloudWatch Logs 會產生費用。

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

註冊 Amazon S3 位置

註冊 Amazon Simple Storage Service AWS Identity and Access Management (Amazon S3) 位置時，您必須指定 (IAM) 角色。Lake Formation 在將臨時登入資料授予存取該位置中資料的整合 AWS 服務時，會擔任該角色。

> **Important**
>
> 避免註冊已啟用申請者付款的 Amazon S3 儲存貯體。對於向 Lake Formation 註冊的儲存貯體，用於註冊儲存貯體的角色一律會被視為申請者。如果儲存貯體是由另一個 AWS 帳戶存取，則如果角色屬於與儲存貯體擁有者相同的帳戶，則會向儲存貯體擁有者收取資料存取費用。

您可以使用 AWS Lake Formation 主控台、Lake Formation API 或 AWS Command Line Interface (AWS CLI) 來註冊 Amazon S3 位置。

開始之前

檢閱用於註冊位置的角色需求。

註冊位置 （主控台）

> **Important**
>
> 下列程序假設 Amazon S3 位置與 Data Catalog 位於相同的 AWS 帳戶中，且位置中的資料未加密。本章的其他章節涵蓋加密位置的跨帳戶註冊和註冊。

1. 在 https：//https://console.aws.amazon.com/lakeformation/ 開啟 AWS Lake Formation 主控台。以資料湖管理員或具有 IAM `lakeformation:RegisterResource` 許可的使用者身分登入。
2. 在導覽窗格的管理下，選取資料湖位置。

3. 選擇註冊位置，然後選擇瀏覽以選取 Amazon Simple Storage Service (Amazon S3) 路徑。
4. （選用，但強烈建議）選取檢閱位置許可，以檢視所選 Amazon S3 位置中的所有現有資源清單及其許可。

   註冊選取的位置可能會導致 Lake Formation 使用者存取該位置已有的資料。檢視此清單可協助您確保現有資料保持安全。
5. 針對 IAM 角色，選擇`AWSServiceRoleForLakeFormationDataAccess`服務連結角色（預設）或符合 中需求的自訂 IAM 角色the section called “用於註冊位置的角色需求”。

   只有在使用自訂 IAM 角色註冊時，才能更新已註冊的位置或其他詳細資訊。若要編輯使用服務連結角色註冊的位置，您應該取消註冊該位置並重新註冊。
6. 選擇啟用資料目錄聯合選項，以允許 Lake Formation 擔任角色並將臨時憑證提供給整合 AWS 服務，以存取聯合資料庫下的資料表。如果某個位置已向 Lake Formation 註冊，而且您想要對聯合資料庫下的資料表使用相同的位置，則需要使用啟用資料目錄聯合選項註冊相同的位置。
7. 選擇混合存取模式，預設不會啟用 Lake Formation 許可。當您在混合存取模式中註冊 Amazon S3 位置時，您可以選擇該位置下資料庫和資料表的主體，以啟用 Lake Formation 許可。

   如需設定混合存取模式的詳細資訊，請參閱 混合存取模式。
8. 選取註冊位置。

註冊位置 (AWS CLI)

1. 向 Lake Formation 註冊新位置

   此範例使用服務連結角色來註冊位置。您可以改為使用 `--role-arn`引數來提供您自己的角色。

   將 *<s3-path>* 取代為有效的 Amazon S3 路徑、將帳號取代為有效的 AWS 帳戶，並將 *<s3-access-role>* 取代為具有註冊資料位置許可的 IAM 角色。

   > **Note**
   >
   > 如果已使用服務連結角色註冊，則無法編輯已註冊位置的屬性。

```
aws lakeformation register-resource \
 --resource-arn arn:aws:s3:::<s3-path> \
 --use-service-linked-role
```

下列範例使用自訂角色來註冊位置。

```
aws lakeformation register-resource \
 --resource-arn arn:aws:s3:::<s3-path> \
 --role-arn arn:aws:iam::<123456789012>:role/<s3-access-role>
```

2. 更新向 Lake Formation 註冊的位置

只有在使用自訂 IAM 角色註冊時，才能編輯已註冊的位置。對於向服務連結角色註冊的位置，您應該取消註冊該位置並重新註冊。如需詳細資訊，請參閱the section called “取消註冊 Amazon S3 位置”。

```
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. 向聯合註冊處於混合存取模式的資料位置

```
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
```

如需詳細資訊，請參閱 RegisterResource API 操作。

> **Note**
>
> 註冊 Amazon S3 位置後，指向位置（或其任何子位置）的任何 AWS Glue 資料表都會傳回 `IsRegisteredWithLakeFormation` 參數的值，如 `GetTable`呼叫`true`中所示。Data Catalog API 操作有已知的限制，例如 `GetTables`和 `SearchTables` 不會更新 `IsRegisteredWithLakeFormation` 參數的值，並傳回預設值，即 false。建議使用 `GetTable` API 來檢視 `IsRegisteredWithLakeFormation` 參數的正確值。

註冊加密的 Amazon S3 位置

Lake Formation 與 AWS Key Management Service(AWS KMS) 整合，可讓您更輕鬆地設定其他整合服務，以加密和解密 Amazon Simple Storage Service (Amazon S3) 位置中的資料。

AWS 受管金鑰 支援客戶受管 AWS KMS keys 和 。目前，只有 Athena 支援用戶端加密/解密。

註冊 Amazon S3 位置時，您必須指定 AWS Identity and Access Management (IAM) 角色。對於加密的 Amazon S3 位置，角色必須具有使用 加密和解密資料的許可 AWS KMS key，或者 KMS 金鑰政策必須將金鑰的許可授予角色。

> **Important**
>
> 避免註冊已啟用申請者付款的 Amazon S3 儲存貯體。對於向 Lake Formation 註冊的儲存貯體，用於註冊儲存貯體的角色一律會被視為申請者。如果儲存貯體是由另一個 AWS 帳戶存取，則如果角色屬於與儲存貯體擁有者相同的帳戶，則會向儲存貯體擁有者收取資料存取費用。

Lake Formation 使用服務連結角色來註冊您的資料位置。不過，此角色有數個限制。由於這些限制，我們建議建立和使用自訂 IAM 角色，而不是為了提高靈活性和控制能力。您為註冊位置而建立的自訂角色必須符合 中指定的要求the section called “用於註冊位置的角色需求”。

> **Important**
>
> 如果您使用 AWS 受管金鑰 來加密 Amazon S3 位置，則無法使用 Lake Formation 服務連結角色。您必須使用自訂角色，並將金鑰上的 IAM 許可新增至角色。本節稍後將提供詳細資訊。

下列程序說明如何註冊使用客戶受管金鑰或 加密的 Amazon S3 位置 AWS 受管金鑰。

- 註冊使用客戶受管金鑰加密的位置
- 使用 註冊加密的位置 AWS 受管金鑰

開始之前

檢閱用於註冊位置的角色需求。

註冊使用客戶受管金鑰加密的 Amazon S3 位置

> **Note**
>
> 如果 KMS 金鑰或 Amazon S3 位置不在與 Data Catalog 相同的 AWS 帳戶中，the section called “跨 AWS 帳戶註冊加密的 Amazon S3 位置”請改為遵循 中的指示。

1. 在 https://console.aws.amazon.com/kms：// 開啟 AWS KMS 主控台，並以 AWS Identity and Access Management (IAM) 管理使用者或可修改用於加密位置之 KMS 金鑰的金鑰政策的使用者身分登入。
2. 在導覽窗格中，選擇客戶受管金鑰，然後選擇所需 KMS 金鑰的名稱。
3. 在 KMS 金鑰詳細資訊頁面上，選擇金鑰政策索引標籤，然後執行下列其中一項操作，將自訂角色或 Lake Formation 服務連結角色新增為 KMS 金鑰使用者：

   - 如果顯示預設檢視 （使用金鑰管理員、金鑰刪除、金鑰使用者和其他 AWS 帳戶區段） – 在金鑰使用者區段下，新增您的自訂角色或 Lake Formation 服務連結角色 `AWSServiceRoleForLakeFormationDataAccess`。
   - 如果顯示金鑰政策 (JSON) – 編輯政策，將自訂角色或 Lake Formation 服務連結角色新增至`AWSServiceRoleForLakeFormationDataAccess`物件「允許使用金鑰」，如下列範例所示。

   > **Note**
   >
   > 如果該物件遺失，請使用範例所示的許可新增物件。此範例使用 服務連結角色。

```
        ...
```

```
        {
            "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. 在 https：//https://console.aws.amazon.com/lakeformation/ 開啟 AWS Lake Formation 主控台。使用 IAM `lakeformation:RegisterResource` 許可以資料湖管理員或使用者身分登入。
5. 在導覽窗格的管理下，選擇資料湖位置。
6. 選擇註冊位置，然後選擇瀏覽以選取 Amazon Simple Storage Service (Amazon S3) 路徑。
7. （選用，但強烈建議）選擇檢閱位置許可，以檢視所選 Amazon S3 位置中的所有現有資源清單及其許可。

   註冊選取的位置可能會導致 Lake Formation 使用者存取該位置已有的資料。檢視此清單可協助您確保現有資料保持安全。
8. 針對 IAM 角色，選擇符合 `AWSServiceRoleForLakeFormationDataAccess`的服務連結角色（預設）或您的自訂角色the section called “用於註冊位置的角色需求”。
9. 選擇註冊位置。

如需服務連結角色的詳細資訊，請參閱Lake Formation 的服務連結角色許可。

註冊使用 加密的 Amazon S3 位置 AWS 受管金鑰

> **Important**
>
> 如果 Amazon S3 位置不在與 Data Catalog 相同的 AWS 帳戶中，the section called “跨 AWS 帳戶註冊加密的 Amazon S3 位置”請改為遵循 中的指示。

1. 建立用於註冊位置的 IAM 角色。確保其符合 中列出的要求the section called “用於註冊位置的角色需求”。
2. 將下列內嵌政策新增至角色。它會將金鑰的許可授予角色。`Resource` 規格必須指定 的 Amazon Resource Name (ARN) AWS 受管金鑰。您可以從 AWS KMS 主控台取得 ARN。若要取得正確的 ARN，請確定您使用 AWS 受管金鑰 與用來加密位置的 相同的 AWS 帳戶和區域登入 AWS KMS 主控台。

   ```
   {
     "Version": "2012-10-17",
     "Statement": [
       {
         "Effect": "Allow",
         "Action": [
           "kms:Encrypt",
           "kms:Decrypt",
           "kms:ReEncrypt*",
           "kms:GenerateDataKey*",
           "kms:DescribeKey"
         ],
         "Resource": "<AWS #### ARN>"
       }
     ]
   }
   ```

   您可以使用 KMS 金鑰別名，而非金鑰 ID - `arn:aws:kms:region:account-id:key/alias/your-key-alias`

   如需詳細資訊，請參閱《 AWS Key Management Service 開發人員指南》中的 中的別名 AWS KMS一節。
3. 在 https：//https://console.aws.amazon.com/lakeformation/ 開啟 AWS Lake Formation 主控台。使用 IAM `lakeformation:RegisterResource` 許可以資料湖管理員或使用者身分登入。

4. 在導覽窗格的管理下，選擇資料湖位置。
5. 選擇註冊位置，然後選擇瀏覽以選取 Amazon S3 路徑。
6. （選用，但強烈建議）選擇檢閱位置許可，以檢視所選 Amazon S3 位置中的所有現有資源清單及其許可。

   註冊選取的位置可能會導致 Lake Formation 使用者存取該位置已有的資料。檢視此清單可協助您確保現有資料保持安全。
7. 針對 IAM 角色，選擇您在步驟 1 中建立的角色。
8. 選擇註冊位置。

在另一個 AWS 帳戶中註冊 Amazon S3 位置

AWS Lake Formation 可讓您跨 AWS 帳戶註冊 Amazon Simple Storage Service (Amazon S3) 位置。例如，如果 AWS Glue Data Catalog 位於帳戶 A，帳戶 A 中的使用者可以在帳戶 B 中註冊 Amazon S3 儲存貯體。

使用 AWS 帳戶 A 中的 AWS Identity and Access Management (IAM) 角色在帳戶 B 中 AWS 註冊 Amazon S3 儲存貯體需要下列許可：

- 帳戶 A 中的角色必須授予帳戶 B 中儲存貯體的許可。
- 帳戶 B 中的儲存貯體政策必須授予帳戶 A 中角色的存取許可。

> **Important**
>
> 避免註冊已啟用申請者付款的 Amazon S3 儲存貯體。對於向 Lake Formation 註冊的儲存貯體，用於註冊儲存貯體的角色一律會被視為申請者。如果儲存貯體是由另一個 AWS 帳戶存取，則如果角色屬於與儲存貯體擁有者相同的帳戶，則會向儲存貯體擁有者收取資料存取費用。
>
> 您無法使用 Lake Formation 服務連結角色來註冊另一個帳戶中的位置。您必須改用使用者定義的角色。角色必須符合 中的要求the section called “用於註冊位置的角色需求”。如需服務連結角色的詳細資訊，請參閱Lake Formation 的服務連結角色許可。

開始之前

檢閱用於註冊位置的角色需求。

在另一個 AWS 帳戶中註冊位置

> **Note**
>
> 如果位置已加密，the section called “跨 AWS 帳戶註冊加密的 Amazon S3 位置”請改為遵循 中的指示。

下列程序假設帳戶 1111-2222-3333 中包含 Data Catalog 的委託人想要註冊`awsexamplebucket1`帳戶 1234-5678-9012 中的 Amazon S3 儲存貯體 。

1. 在帳戶 1111-2222-3333 中，登入 AWS Management Console 並在 開啟 IAM 主控台https://console.aws.amazon.com/iam/。
2. 建立新的角色或檢視符合 中需求的現有角色the section called “用於註冊位置的角色需求”。確保角色授予 上的 Amazon S3 許可`awsexamplebucket1`。
3. 開啟位於 https://console.aws.amazon.com/s3/ 的 Amazon S3 主控台。使用帳戶 1234-5678-9012 登入。
4. 在儲存貯體名稱清單中，選擇儲存貯體名稱 `awsexamplebucket1`。
5. 選擇許可。
6. 在許可頁面上，選擇儲存貯體政策。
7. 在儲存貯體政策編輯器中，貼上下列政策。將 *<role-name>* 取代為您的角色名稱。

```
{
    "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. 選擇儲存。
9. 在 https：//https://console.aws.amazon.com/lakeformation/ 開啟 AWS Lake Formation 主控台。以資料湖管理員或具有足夠註冊位置許可的使用者身分登入帳戶 1111-2222-3333。
10. 在導覽窗格的管理下，選擇資料湖位置。
11. 在資料湖位置頁面上，選擇註冊位置。
12. 在註冊位置頁面上，針對 Amazon S3 路徑，輸入儲存貯體名稱 `s3://awsexamplebucket1`。

> **Note**
>
> 您必須輸入儲存貯體名稱，因為當您選擇瀏覽時，跨帳戶儲存貯體不會出現在清單中。

13. 針對 IAM 角色，選擇您的角色。
14. 選擇註冊位置。

跨 AWS 帳戶註冊加密的 Amazon S3 位置

AWS Lake Formation 與 AWS Key Management Service(AWS KMS) 整合，可讓您更輕鬆地設定其他整合服務，以加密和解密 Amazon Simple Storage Service (Amazon S3) 位置中的資料。

AWS 受管金鑰 支援客戶受管金鑰和 。不支援用戶端加密/解密。

> **Important**
>
> 避免註冊已啟用申請者付款的 Amazon S3 儲存貯體。對於向 Lake Formation 註冊的儲存貯體，用於註冊儲存貯體的角色一律會被視為申請者。如果儲存貯體是由另一個 AWS 帳戶存取，則如果角色屬於與儲存貯體擁有者相同的帳戶，則會向儲存貯體擁有者收取資料存取費用。

本節說明如何在下列情況下註冊 Amazon S3 位置：

- Amazon S3 位置中的資料會使用在其中建立的 KMS 金鑰進行加密 AWS KMS。
- Amazon S3 位置不在與 相同的 AWS 帳戶中 AWS Glue Data Catalog。
- KMS 金鑰與 Data Catalog 位於或不在相同的 AWS 帳戶中。

在 AWS 帳戶 A 中使用 (IAM) 角色在 AWS 帳戶 B 中註冊 AWS KMS加密的 AWS Identity and Access Management Amazon S3 儲存貯體需要下列許可：

- 帳戶 A 中的角色必須授予帳戶 B 中儲存貯體的許可。
- 帳戶 B 中的儲存貯體政策必須授予帳戶 A 中角色的存取許可。
- 如果 KMS 金鑰位於帳戶 B 中，金鑰政策必須授予帳戶 A 中角色的存取權，而帳戶 A 中的角色必須授予 KMS 金鑰的許可。

在下列程序中，您會在包含 Data Catalog AWS 的帳戶中建立角色（先前討論中的帳戶 A)。然後，您可以使用此角色來註冊位置。Lake Formation 在存取 Amazon S3 中的基礎資料時擔任此角色。擔任的角色具有 KMS 金鑰所需的許可。因此，您不需要將 KMS 金鑰的許可授予使用 ETL 任務或整合服務存取基礎資料的委託人，例如 Amazon Athena。

> **Important**
>
> 您無法使用 Lake Formation 服務連結角色來註冊另一個帳戶中的位置。您必須改用使用者定義的角色。角色必須符合 中的要求the section called “用於註冊位置的角色需求”。如需服務連結角色的詳細資訊，請參閱Lake Formation 的服務連結角色許可。

開始之前

檢閱用於註冊位置的角色需求。

跨 AWS 帳戶註冊加密的 Amazon S3 位置

1. 在與 Data Catalog 相同的 AWS 帳戶中，登入 AWS Management Console 並在 開啟 IAM 主控台https://console.aws.amazon.com/iam/。
2. 建立新的角色或檢視符合 中需求的現有角色the section called “用於註冊位置的角色需求”。確保角色包含授予 Amazon S3 位置許可的政策。

3. 如果 KMS 金鑰不在與 Data Catalog 相同的帳戶中，請將授予 KMS 金鑰必要許可的內嵌政策新增至角色。政策範例如下。將 *<cmk-region>* 和 *<cmk-account-id>* 取代為 KMS 金鑰的區域和帳戶號碼。以金鑰 ID 取代 *<key-id>*。

```
{
    "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. 在 Amazon S3 主控台上，新增儲存貯體政策，將所需的 Amazon S3 許可授予角色。以下為儲存貯體政策的範例。將 *<catalog-account-id>* 取代為 Data Catalog 的 AWS 帳號、將 *<role-name>* 取代為角色名稱，並將 *<bucket-name>* 取代為儲存貯體名稱。

```
{
    "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. 在 中 AWS KMS，將角色新增為 KMS 金鑰的使用者。

   a. 在 https：//https://console.aws.amazon.com/kms 開啟 AWS KMS 主控台。然後，以管理員使用者身分登入，或以可以修改用於加密位置之 KMS 金鑰的金鑰政策的使用者身分登入。

   b. 在導覽窗格中，選擇客戶受管金鑰，然後選擇 KMS 金鑰的名稱。

   c. 在 KMS 金鑰詳細資訊頁面的金鑰政策索引標籤下，如果金鑰政策的 JSON 檢視未顯示，請選擇切換到政策檢視。

   d. 在金鑰政策區段中，選擇編輯，然後將角色的 Amazon Resource Name (ARN) 新增至`Allow use of the key`物件，如下列範例所示。

   > **Note**
   >
   > 如果該物件遺失，請使用範例所示的許可新增物件。

```
        ...
        {
            "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": "*"
```

```
        },
        ...
```

如需詳細資訊，請參閱《 AWS Key Management Service 開發人員指南》中的允許其他帳戶中的使用者使用 KMS 金鑰。

6. 在 https：//https://console.aws.amazon.com/lakeformation/ 開啟 AWS Lake Formation 主控台。以資料湖管理員身分登入 Data Catalog AWS 帳戶。
7. 在導覽窗格的管理下，選擇資料湖位置。
8. 選擇註冊位置。
9. 在註冊位置頁面上，針對 Amazon S3 路徑，輸入位置路徑為 **s3://*<bucket>*/*<prefix>***。將 *<bucket>* 取代為儲存貯體的名稱，並將 *<prefix>* 取代為位置的其餘路徑。

> **Note**
>
> 您必須輸入路徑，因為當您選擇瀏覽時，跨帳戶儲存貯體不會出現在清單中。

10. 針對 IAM 角色，從步驟 2 選擇角色。
11. 選擇註冊位置。

取消註冊 Amazon S3 位置

如果您不想再由 Lake Formation 管理 Amazon Simple Storage Service (Amazon S3) 位置，您可以取消註冊。取消註冊位置不會影響在該位置上授予的 Lake Formation 資料位置許可。您可以重新註冊取消註冊的位置，資料位置許可仍然有效。您可以使用不同的角色來重新註冊位置。

取消註冊位置（主控台）

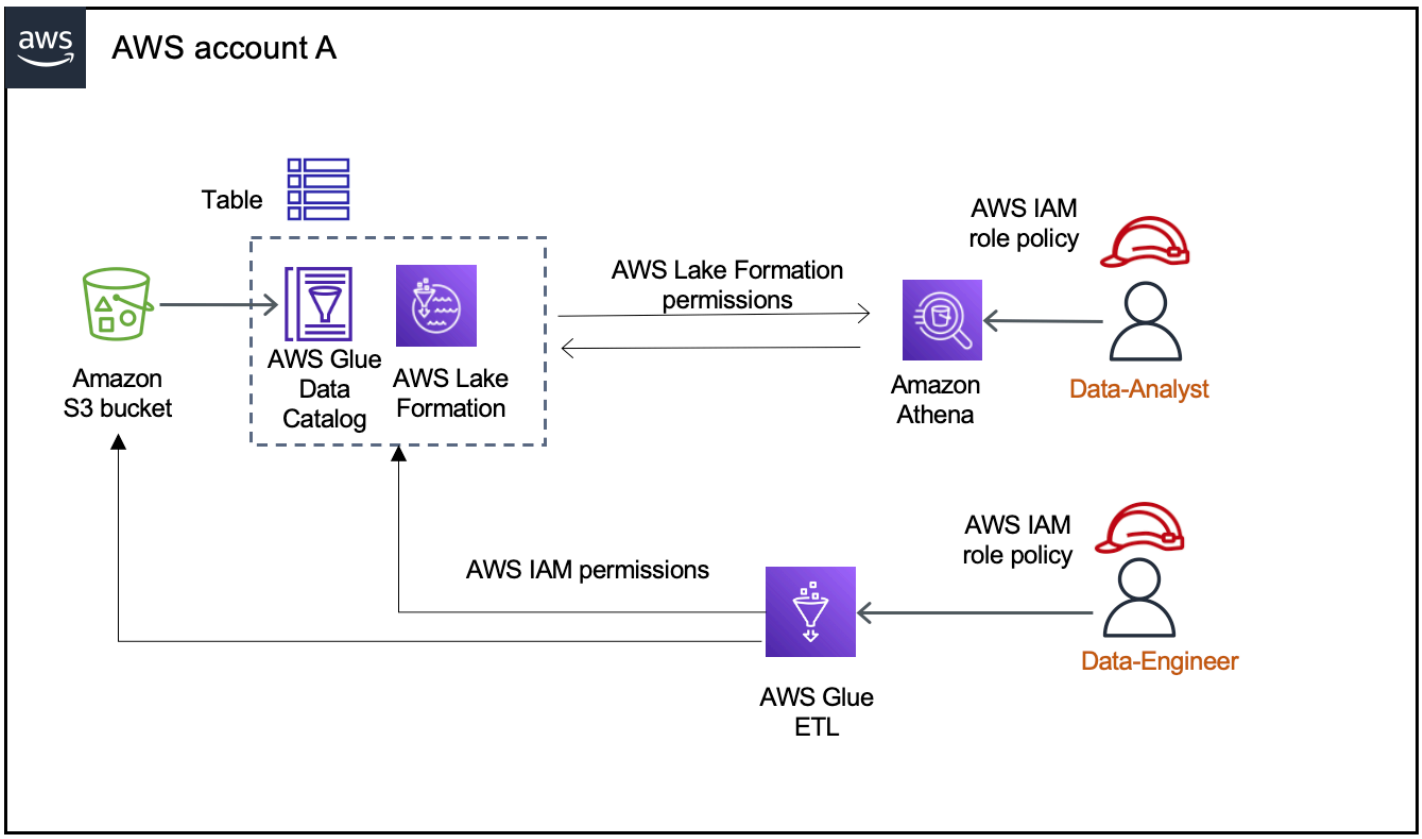
1. 在 https：//https://console.aws.amazon.com/lakeformation/ 開啟 AWS Lake Formation 主控台。以資料湖管理員或具有 IAM `lakeformation:RegisterResource` 許可的使用者身分登入。
2. 在導覽窗格的管理下，選擇資料湖位置。
3. 選取位置，然後在動作功能表上，選擇移除。
4. 出現確認提示時，請選擇移除。

混合存取模式

AWS Lake Formation 混合存取模式支援通往相同 AWS Glue Data Catalog 物件的兩個許可路徑。在第一個路徑中，Lake Formation 可讓您選取特定主體，並授予他們 Lake Formation 透過選擇加入存取目錄、資料庫、資料表和檢視的許可。第二個路徑允許所有其他主體透過 Amazon S3 和 AWS Glue 動作的預設 IAM 主體政策存取這些資源。

向 Lake Formation 註冊 Amazon S3 位置時，您可以選擇針對此位置的所有資源強制執行 Lake Formation 許可，或使用混合存取模式。根據預設`CREATE_TABLE`，混合存取模式只會強制執行 、`CREATE_PARTITION`、 `UPDATE_TABLE`許可。當 Amazon S3 位置處於混合模式時，您可以透過選擇該位置下資料目錄物件的主體來啟用 Lake Formation 許可。
這表示 Lake Formation 許可和 IAM 許可都可以控制對該資料的存取。這表示選擇加入主體需要 Lake Formation 許可和 IAM 許可才能存取資料，而non-opted-in主體將繼續僅使用 IAM 許可存取資料。

因此，混合存取模式可讓您靈活地為特定一組使用者選擇性地為 Data Catalog 中的目錄、資料庫和資料表啟用 Lake Formation，而不會中斷其他現有使用者或工作負載的存取。



如需注意事項和限制，請參閱 混合存取模式的考量和限制 。

術語和定義

以下是 Data Catalog 資源的定義，取決於您如何設定存取許可：

Lake Formation 資源

向 Lake Formation 註冊的資源。使用者需要 Lake Formation 許可才能存取資源。

AWS Glue 資源

未向 Lake Formation 註冊的資源。使用者只需要 IAM 許可即可存取資源，因為它具有`IAMAllowedPrincipals`群組許可。Lake Formation 許可不會強制執行。

如需`IAMAllowedPrincipals`群組許可的詳細資訊，請參閱 中繼資料許可。

混合資源

在混合存取模式中註冊的資源。根據存取資源的使用者，資源會動態切換為 Lake Formation 資源或 AWS Glue 資源。

常見的混合存取模式使用案例

您可以使用混合存取模式，在單一帳戶和跨帳戶資料共用案例中提供存取權：

單一帳戶案例

- 將 AWS Glue 資源轉換為混合資源 – 在此案例中，您目前並未使用 Lake Formation，但想要為 Data Catalog 物件採用 Lake Formation 許可。當您以混合存取模式註冊 Amazon S3 位置時，您可以將 Lake Formation 許可授予選擇加入指向該位置之特定資料庫和資料表的使用者。
- 將 Lake Formation 資源轉換為混合資源 – 目前，您正在使用 Lake Formation 許可來控制 Data Catalog 資料庫的存取權，但想要使用 Amazon S3 的 IAM 許可提供新委託人的存取權，而不會 AWS Glue 中斷現有的 Lake Formation 許可。

  當您將資料位置註冊更新為混合存取模式時，新的主體可以使用 IAM 許可政策存取指向 Amazon S3 位置的資料目錄資料庫，而不會中斷現有使用者的 Lake Formation 許可。

  在更新資料位置註冊以啟用混合存取模式之前，您必須先選擇使用 Lake Formation 許可存取資源的主體。
  這是為了防止目前工作流程的潛在中斷。
  您也需要將資料庫中資料表的`Super`許可授予 `IAMAllowedPrincipal`群組。

跨帳戶資料共用案例

- 使用混合存取模式共用 AWS Glue 資源 – 在此案例中，生產者帳戶在資料庫中具有資料表，目前使用 Amazon S3 AWS Glue 的 IAM 許可政策和動作與取用者帳戶共用。資料庫的資料位置未向 Lake Formation 註冊。

  在混合存取模式中註冊資料位置之前，您需要將跨帳戶版本設定更新為第 4 版。當`IAMAllowedPrincipal`群組具有資源的 AWS RAM 許可時，第 4 版提供跨帳戶共用所需的新`Super`許可政策。對於具有`IAMAllowedPrincipal`群組許可的資源，您可以將 Lake Formation 許可授予外部帳戶，並選擇加入以使用 Lake Formation 許可。收件人帳戶中的資料湖管理員可以將 Lake Formation 許可授予帳戶中的主體，並選擇加入以強制執行 Lake Formation 許可。
- 使用混合存取模式共用 Lake Formation 資源 – 目前，生產者帳戶在資料庫中具有與強制執行 Lake Formation 許可的取用者帳戶共用的資料表。資料庫的資料位置已向 Lake Formation 註冊。

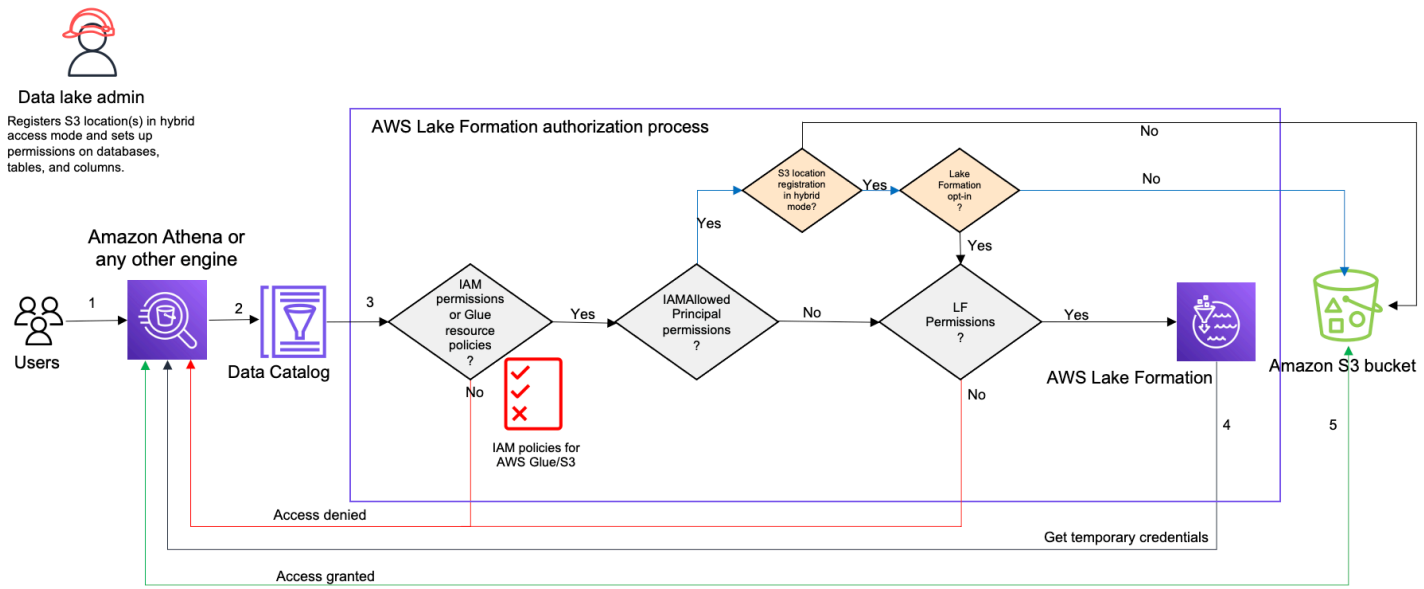
  在此情況下，您可以將 Amazon S3 位置註冊更新為混合存取模式，並使用 Amazon S3 儲存貯體政策和 Data Catalog 資源政策，將來自 Amazon S3 的資料和來自 Data Catalog 的中繼資料分享給取用者帳戶中的主體。您需要重新授予現有的 Lake Formation 許可，並在更新 Amazon S3 位置註冊之前選擇加入委託人。此外，您需要將資料庫中資料表的`Super`許可授予`IAMAllowedPrincipals`群組。

主題

- 混合存取模式的運作方式
- 設定混合存取模式 - 常見案例
- 從混合存取模式移除主體和資源
- 在混合存取模式中檢視主體和資源
- 其他資源

混合存取模式的運作方式

下圖顯示當您查詢 Data Catalog 資源時，Lake Formation 授權如何在混合存取模式中運作。



在存取資料湖中的資料之前，具有管理許可的資料湖管理員或使用者會設定個別 Data Catalog 資料表使用者政策，以允許或拒絕存取 Data Catalog 中的資料表。然後，具有執行`RegisterResource`操作許可的主體會將資料表的 Amazon S3 位置向 Lake Formation 註冊為混合存取模式。如果資料位置未向 Lake Formation 註冊，管理員會將 Lake Formation 許可授予 Data Catalog 資料庫和資料表上的特定使用者，並選擇讓他們在混合存取模式中使用這些資料庫和資料表的 Lake Formation 許可。

1. 提交查詢 - 委託人使用整合的服務提交查詢或 ETL 指令碼，例如 Amazon Athena AWS Glue、Amazon EMR 或 Amazon Redshift Spectrum。
2. 請求資料 - 整合式分析引擎會識別正在請求的資料表，並將中繼資料請求傳送至 Data Catalog (`GetTable`、`GetDatabase`)。
3. 檢查許可 - Data Catalog 會使用 Lake Formation 驗證查詢主體的存取許可。
   a. 如果資料表未連接`IAMAllowedPrincipals`群組許可，則會強制執行 Lake Formation 許可。
   b. 如果委託人已選擇在混合存取模式中使用 Lake Formation 許可，且資料表已連接`IAMAllowedPrincipals`群組許可，則會強制執行 Lake Formation 許可。查詢引擎會套用從 Lake Formation 收到的篩選條件，並將資料傳回給使用者。
   c. 如果資料表位置未向 Lake Formation 註冊，且委託人尚未選擇在混合存取模式中使用 Lake Formation 許可，則 Data Catalog 會檢查資料表是否已連接`IAMAllowedPrincipals`群組許可。如果此許可存在於資料表中，則帳戶中的所有主體都會取得資料表上的 `Super`或 `All`許可。

   除非資料位置已向 Lake Formation 註冊，否則即使選擇加入，Lake Formation 登入資料販賣也無法使用。

4. 取得憑證 – Data Catalog 會檢查並告知引擎資料表位置是否已向 Lake Formation 註冊。如果基礎資料已向 Lake Formation 註冊，分析引擎會請求 Lake Formation 提供臨時憑證，以存取 Amazon S3 儲存貯體中的資料。
5. 取得資料 – 如果委託人有權存取資料表資料，Lake Formation 會提供整合式分析引擎的暫時存取權。使用臨時存取，分析引擎會從 Amazon S3 擷取資料，並執行必要的篩選，例如資料欄、資料列或儲存格篩選。當引擎完成執行任務時，會將結果傳回給使用者。此程序稱為憑證販售程序。如需詳細資訊，
   請參閱 與 Lake Formation 整合。
6. 如果資料表的資料位置未向 Lake Formation 註冊，則分析引擎的第二個呼叫會直接對 Amazon S3 進行。評估相關的 Amazon S3 儲存貯體政策和 IAM 使用者政策以進行資料存取。每當您使用 IAM 政策時，請務必遵循 IAM 最佳實務。如需詳細資訊，請參閱《IAM 使用者指南》中的 IAM 中的安全最佳實務。

設定混合存取模式 - 常見案例

與 Lake Formation 許可一樣，您通常有兩種類型的案例，您可以在其中使用混合存取模式來管理資料存取：提供存取給一個 中的委託人， AWS 帳戶 並提供外部 AWS 帳戶 或委託人的存取。

本節提供在下列情況下設定混合存取模式的說明：

在一個 內以混合存取模式管理許可 AWS 帳戶

- 將 AWS Glue 資源轉換為混合資源 – 您目前使用 Amazon S3 的 IAM 許可，為帳戶中所有主體提供資料庫中的資料表存取權， AWS Glue 但想要採用 Lake Formation 來逐步管理許可。

- 將 Lake Formation 資源轉換為混合資源 – 您目前使用 Lake Formation 來管理您帳戶中所有主體資料庫中資料表的存取權，但只想將 Lake Formation 用於特定主體。您想要在相同的資料庫和資料表上使用 AWS Glue 和 Amazon S3 的 IAM 許可，為新主體提供存取權。

跨 在混合存取模式中管理許可 AWS 帳戶

- 使用混合存取模式共用 AWS Glue 資源 – 您目前不是使用 Lake Formation 來管理資料表的許可，但想要套用 Lake Formation 許可來為另一個帳戶中的主體提供存取權。
- 使用混合存取模式共用 Lake Formation 資源 – 您正在使用 Lake Formation 來管理資料表的存取權，但想要在相同資料庫 AWS Glue 和資料表上使用 和 Amazon S3 的 IAM 許可，為另一個帳戶中的主體提供存取權。

設定混合存取模式 – 高階步驟

1. 選取混合存取模式，向 Lake Formation 註冊 Amazon S3 資料位置。
2. 主體必須擁有資料湖位置的DATA_LOCATION許可，才能建立指向該位置的資料目錄資料表或資料庫。
3. 將跨帳戶版本設定設為第 4 版。
4. 將精細許可授予資料庫和資料表上的特定 IAM 使用者或角色。同時，請務必為資料庫上的IAMAllowedPrincipals群組和資料庫中所有Super或選取的資料表設定 或 All許可。
5. 選擇主體和資源。帳戶中的其他主體可以使用 和 Amazon S3 動作的 IAM 許可政策繼續存取資料庫 AWS Glue 和資料表。
6. 選擇性地為選擇使用 Lake Formation 許可的主體清除 Amazon S3 的 IAM 許可政策。

設定混合存取模式的先決條件

以下是設定混合存取模式的先決條件：

> **Note**
>
> 我們建議 Lake Formation 管理員以混合存取模式註冊 Amazon S3 位置，並選擇加入主體和資源。

1. 授予資料位置許可 (DATA_LOCATION_ACCESS)，以建立指向 Amazon S3 位置的 Data Catalog 資源。資料位置許可控制建立指向特定 Amazon S3 位置之 Data Catalog 目錄、資料庫和資料表的能力。
2. 若要在混合存取模式中與其他帳戶共用 Data Catalog 資源 （ 而不從資源移除IAMAllowedPrincipals群組許可），您需要將跨帳戶版本設定更新為第 4 版。若要使用 Lake Formation 主控台更新版本，請在資料目錄設定頁面的跨帳戶版本設定下選擇第 4 版。

   您也可以使用 put-data-lake-settings AWS CLI 命令將 CROSS_ACCOUNT_VERSION 參數設定為第 4 版：

```
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. 若要在混合存取模式中授予跨帳戶許可，授予者必須擁有 AWS Glue 和 AWS RAM 服務所需的 IAM 許可。 AWS 受管政策會AWSLakeFormationCrossAccountManager授予必要的許可。為了在混合存取模式中啟用跨帳戶資料共用，我們新增了兩個新的 IAM 許可來更新 AWSLakeFormationCrossAccountManager受管政策：

   - ram:ListResourceSharePermissions
   - ram:AssociateResourceSharePermission

   > **Note**
   >
   > 如果您不是使用授予者角色的 AWS 受管政策，請將上述政策新增至您的自訂政策。

Amazon S3 儲存貯體位置和使用者存取

當您在 中建立目錄、資料庫或資料表時 AWS Glue Data Catalog，您可以指定基礎資料的 Amazon S3 儲存貯體位置，並將其註冊至 Lake Formation。下表說明根據資料表或資料庫的 Amazon S3 資料位置，許可如何適用於 AWS Glue 和 Lake Formation 使用者 （委託人）。

向 Lake Formation 註冊的 Amazon S3 位置

資料庫的 Amazon S3 位置	AWS Glue 使用者	Lake Formation 使用者
向 Lake Formation 註冊 （混合存取模式或 Lake Formation 模式）	透過繼承 IAMAllowedPrincipals 群組 （超級存取） 許可，擁有 Amazon S3 資料位置的讀取/寫入存取權。	繼承從其授予的 CREATE TABLE 許可建立資料表的許可。

資料庫的 Amazon S3 位置	AWS Glue 使用者	Lake Formation 使用者
沒有相關聯的 Amazon S3 位置	執行 CREATE TABLE 和 INSERT TABLE 陳述式需要明確的 DATA LOCATION 許可。	執行 CREATE TABLE 和 INSERT TABLE 陳述式需要明確的 DATA LOCATION 許可。

IsRegisteredWithLakeFormation 資料表屬性

資料表的 `IsRegisteredWithLakeFormation` 屬性指出資料表的資料位置是否已向請求者的 Lake Formation 註冊。如果位置的許可模式已註冊為 Lake Formation，則 `IsRegisteredWithLakeFormation` 屬性`true`適用於存取資料位置的所有使用者，因為所有使用者都被視為已選擇加入該資料表。如果位置是在混合存取模式中註冊，則 值`true`只會針對已選擇加入該資料表的使用者設為 。

`IsRegisteredWithLakeFormation` 的運作方式

許可模式	使用者/角色	**`IsRegisteredWithLakeFormation`**	描述
Lake Formation	全部	True	向 Lake Formation 註冊位置時，所有使用者的 `IsRegisteredWithLakeFormation` 屬性都會設為 true。這表示 Lake Formation 中定義的許可適用於註冊的位置。登入資料販賣將由 Lake Formation 完成。
混合存取模式	選擇加入	True	對於選擇使用 Lake Formation 進行資料表資料存取和管理的使用者，該資料表`true`的 `IsRegiste`

許可模式	使用者/角色	**IsRegisteredWithLakeFormation**	描述
			redWithLakeFormation 屬性將設定為 。它們受註冊位置 Lake Formation 中定義的許可政策約束。
混合存取模式	未選擇加入	False	對於尚未選擇使用 Lake Formation 許可的使用者，IsRegisteredWithLakeFormation 屬性會設定為 false。它們不受 Lake Formation 中為註冊位置定義的許可政策約束。反之，使用者將遵循 Amazon S3 許可政策。

將 AWS Glue 資源轉換為混合資源

請依照下列步驟，在混合存取模式中註冊 Amazon S3 位置，並在不中斷現有 Data Catalog 使用者的資料存取的情況下加入新的 Lake Formation 使用者。

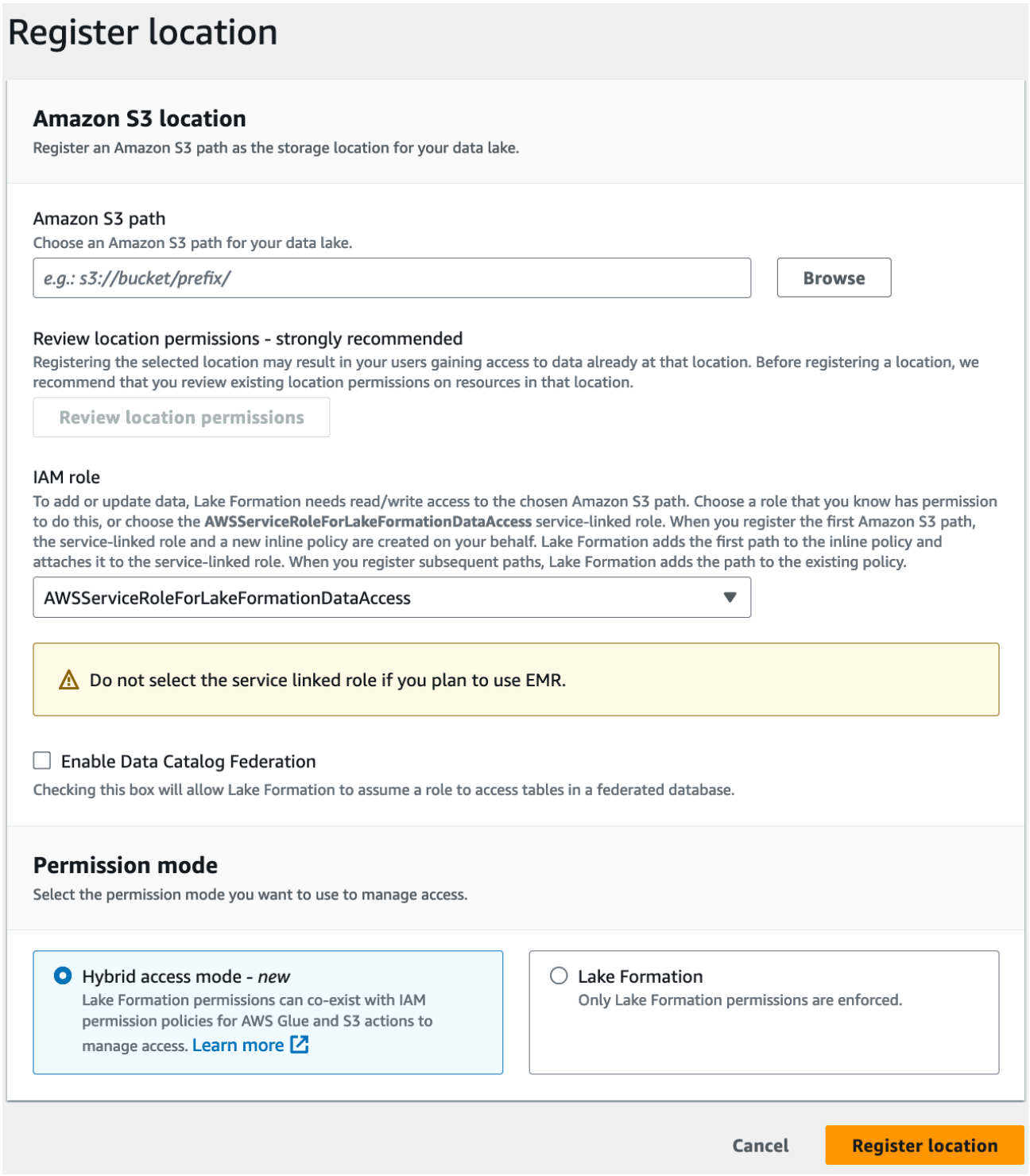
案例描述 - 資料位置未向 Lake Formation 註冊，使用者對 Data Catalog 資料庫和資料表的存取取決於 Amazon S3 和 AWS Glue 動作的 IAM 許可政策。
根據預設， IAMAllowedPrincipals群組具有資料庫中所有資料表的Super許可。

為未向 Lake Formation 註冊的資料位置啟用混合存取模式

1. 註冊啟用混合存取模式的 Amazon S3 位置。

   Console

   1. 以資料湖管理員身分登入 Lake Formation 主控台。
   2. 在導覽窗格中，選擇管理下的資料湖位置。
   3. 選擇註冊位置。



4. 在註冊位置視窗中，選擇您要向 Lake Formation 註冊的 Amazon S3 路徑。
5. 針對 IAM 角色，選擇`AWSServiceRoleForLakeFormationDataAccess`服務連結角色（預設）或自訂 IAM
   符合 中需求的 角色[用於註冊位置的角色需求](#)。

6. 選擇混合存取模式，將精細的 Lake Formation 存取控制政策套用至指向註冊位置的選擇加入主體和 Data Catalog 資料庫和資料表。

   選擇 Lake Formation 以允許 Lake Formation 授權對註冊位置的存取請求。

7. 選擇註冊位置。

AWS CLI

以下是使用 HybridAccessEnabled：true/false 向 Lake Formation 註冊資料位置的範例。HybridAccessEnabled 參數的預設值為 false。以有效值取代 Amazon S3 路徑、角色名稱和 AWS 帳戶 ID。

```
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. 授予許可並選擇加入主體，以對混合存取模式中的資源使用 Lake Formation 許可

在混合存取模式中選擇加入主體和資源之前，請確認 Super或 IAMAllowedPrincipals群組的All許可存在於已在混合存取模式中向 Lake Formation 註冊位置的資料庫和資料表上。

> **Note**
>
> 您無法在資料庫中授予 All tables的IAMAllowedPrincipals群組許可。您需要從下拉式選單中分別選取每個資料表，並授予許可。此外，當您在資料庫中建立新資料表時，您可以選擇在資料目錄設定Use only IAM access control for new tables in new databases中使用 。當您在資料庫中建立新資料表時，此選項會自動將Super許可授予 IAMAllowedPrincipals群組。

Console

1. 在 Lake Formation 主控台的資料目錄下，選擇目錄、資料庫或資料表。
2. 從清單中選擇目錄、資料庫或資料表，然後從動作功能表中選擇授予。

3. 選擇主體，以使用具名資源方法或 LF 標籤授予資料庫、資料表和資料欄的許可。

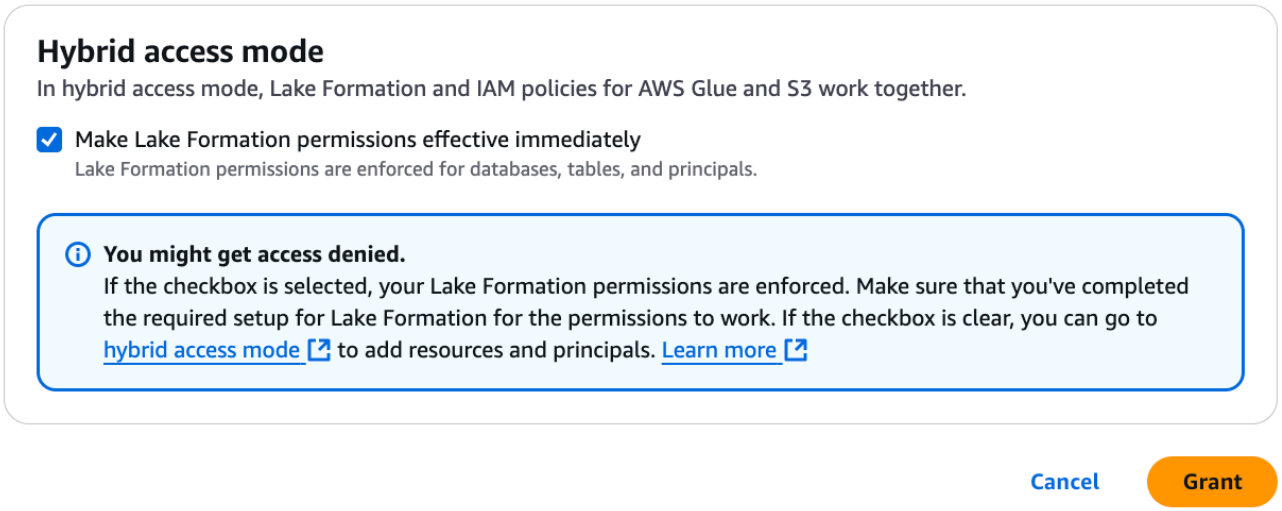
   或者，選擇資料許可，從清單中選擇要授予許可的委託人，然後選擇授予。

   如需授予資料許可的詳細資訊，請參閱 授予 Data Catalog 資源的許可。

   > **Note**
   >
   > 如果您要授予委託人建立資料表許可，您也需要將資料位置許可 (DATA_LOCATION_ACCESS) 授予委託人。更新資料表不需要此許可。如需詳細資訊，請參閱授予資料位置許可。

4. 當您使用具名資源方法來授予許可時，在授予資料許可頁面的下一節提供選擇加入委託人和資源的選項。

   選擇讓 Lake Formation 許可立即生效，以啟用委託人和資源的 Lake Formation 許可。

5. 選擇 Grant (授予)。

   當您在指向資料位置的資料表 A 上選擇加入委託人 A 時，如果資料位置註冊為混合模式，則允許委託人 A 使用 Lake Formation 許可來存取此資料表的位置。

AWS CLI

以下是在混合存取模式中選擇加入委託人和資料表的範例。將角色名稱、 AWS 帳戶 ID、資料庫名稱和資料表名稱取代為有效值。

```
aws lakeformation create-lake-formation-opt-in --cli-input-json file://file path
```

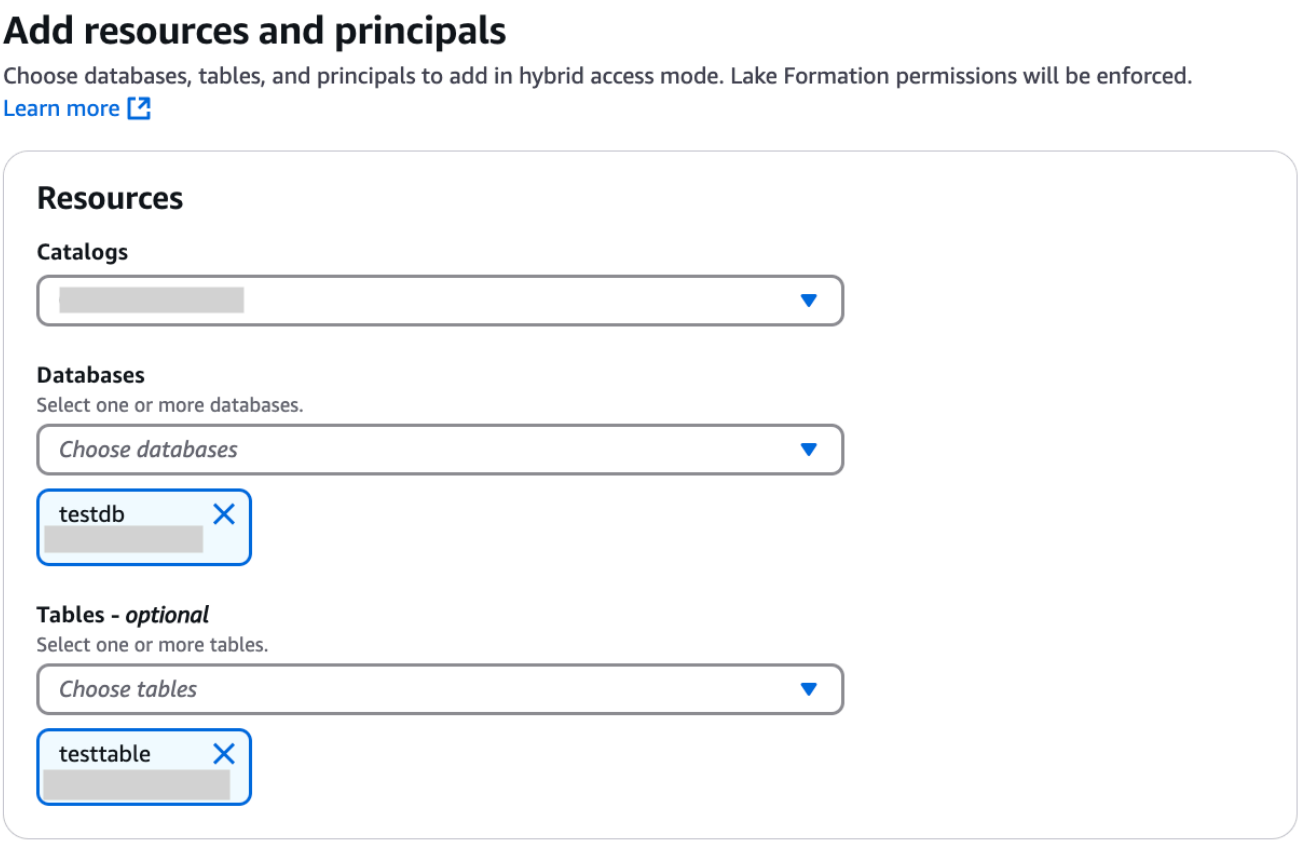
```
json:
  {
        "Principal": {
            "DataLakePrincipalIdentifier":
 "arn:aws:iam::<123456789012>:role/<hybrid-access-role>"
        },
        "Resource": {
            "Table": {
                "CatalogId": "<123456789012>",
                "DatabaseName": "<hybrid_test>",
                "Name": "<hybrid_test_table>"
            }
        }
    }
```

a. (Optional) 如果您選擇 LF-Tags 授予許可，您可以選擇讓主體在不同的步驟中使用 Lake Formation 許可。您可以在左側導覽列的許可下選擇混合存取模式來執行此操作。

b. 在混合存取模式頁面的下區段中，選擇新增以將資源和主體新增至混合存取模式。

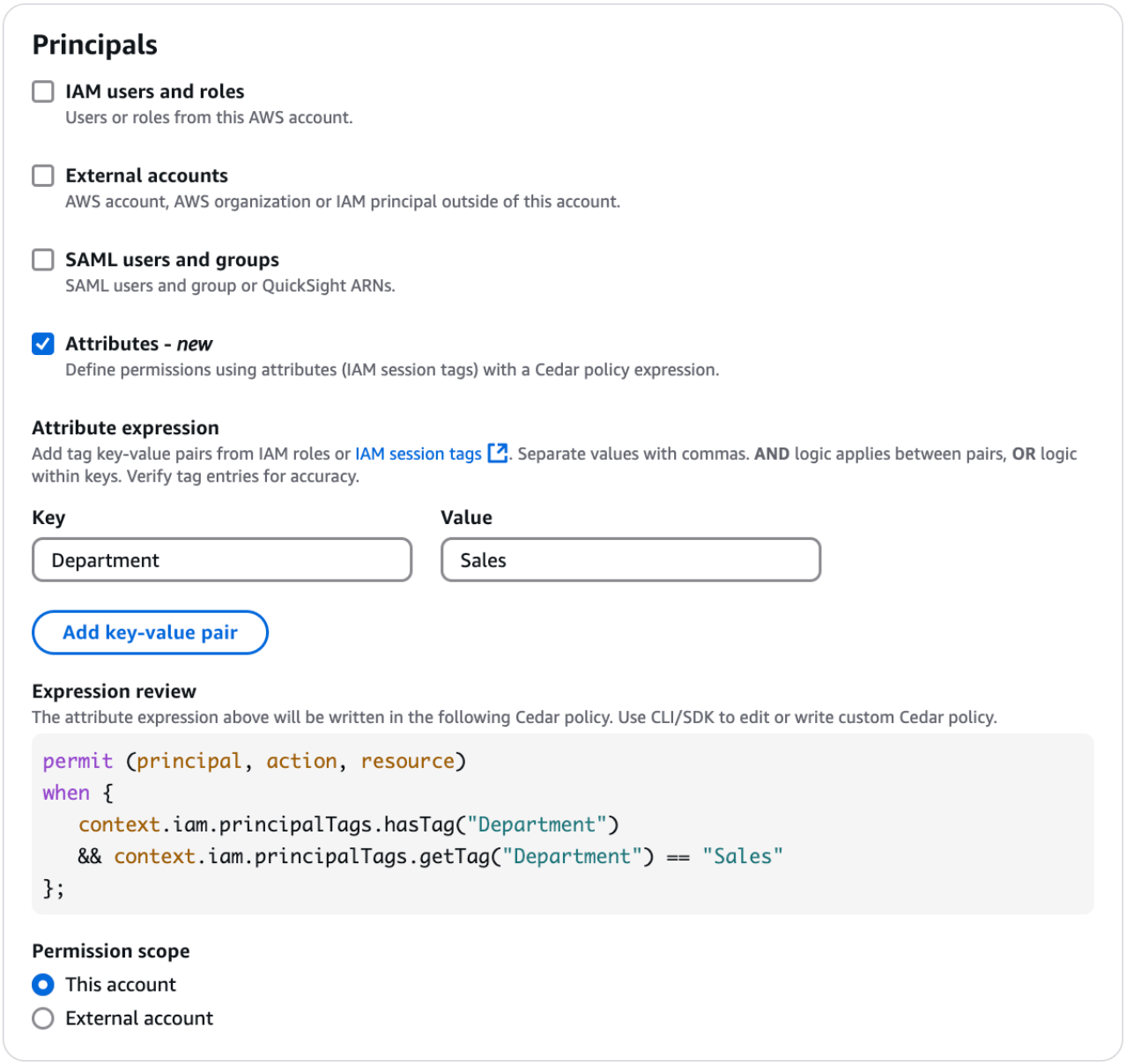
c. 在新增資源和主體頁面上，選擇在混合存取模式中註冊的目錄、資料庫和資料表。

   您可以在資料庫All tables下選擇 以授予存取權。



d. 選擇主體選擇在混合存取模式中使用 Lake Formation 許可。

- 委託人 – 您可以在相同帳戶或其他帳戶中選擇 IAM 使用者和角色。您也可以選擇 SAML 使用者和群組。
- 屬性 – 根據屬性選取要授予許可的屬性。



- 輸入鍵/值對，以根據屬性建立授予。在主控台上檢閱 Cedar 政策表達式。如需 Cedar 的詳細資訊，請參閱什麼是 Cedar？| Cedar 政策語言參考GuideLink。
- 選擇新增。

  具有相符屬性的所有 IAM 角色/使用者都會獲得存取權。

e. 選擇新增。

將 Lake Formation 資源轉換為混合資源

如果您目前正在為 Data Catalog 資料庫和資料表使用 Lake Formation 許可，您可以編輯位置註冊屬性以啟用混合存取模式。這可讓您使用 Amazon S3 的 IAM 許可政策和 AWS Glue 動作，為新主體提供對相同資源的存取權，而不會中斷現有的 Lake Formation 許可。

案例描述 - 下列步驟假設您已向 Lake Formation 註冊資料位置，而且您已為指向該位置的資料庫、資料表或資料欄上的主體設定許可。如果位置已向服務連結角色註冊，則無法更新位置參數並啟用混合存取模式。根據預設，`IAMAllowedPrincipals`群組對資料庫及其所有資料表具有超級許可。

> **Important**
>
> 請勿在未選擇在此位置存取資料的主體的情況下，將位置註冊更新為混合存取模式。

為向 Lake Formation 註冊的資料位置啟用混合存取模式

1.

   > **Warning**
   >
   > 我們不建議將 Lake Formation 受管資料位置轉換為混合存取模式，以避免中斷其他現有使用者或工作負載的許可政策。

   選擇具有 Lake Formation 許可的現有委託人。

   1. 列出並檢閱您已授予目錄、資料庫和資料表主體的許可。如需詳細資訊，請參閱在 Lake Formation 中檢視資料庫和資料表許可。
   2. 在左側導覽列的許可下選擇混合存取模式，然後選擇新增。
   3. 在新增主體和資源頁面上，從您要在混合存取模式中使用的 Amazon S3 資料位置中選擇目錄、資料庫和資料表。選擇已具有 Lake Formation 許可的主體。
   4. 選擇新增以選擇加入主體，以在混合存取模式中使用 Lake Formation 許可。
2. 選擇混合存取模式選項，更新 Amazon S3 儲存貯體/字首註冊。

   Console

   1. 以資料湖管理員身分登入 Lake Formation 主控台。
   2. 在導覽窗格的註冊和擷取下，選擇資料湖位置。
   3. 選取位置，然後在動作功能表上，選擇編輯。

4. 選擇混合存取模式。
5. 選擇儲存。
6. 在資料目錄下，選取資料庫或資料表，並將 Super或 All許可授予名為 的虛擬群組IAMAllowedPrincipals。
7. 當您更新位置註冊屬性時，請確認您現有的 Lake Formation 使用者的存取權未中斷。以 Lake Formation 主體身分登入 Athena 主控台，並在指向更新位置的資料表上執行範例查詢。

   同樣地，驗證使用 IAM 許可政策來存取資料庫和資料表之 AWS Glue 使用者的存取權。

AWS CLI

以下是使用 HybridAccessEnabled：true/false 向 Lake Formation 註冊資料位置的範例。HybridAccessEnabled 參數的預設值為 false。以有效值取代 Amazon S3 路徑、角色名稱和 AWS 帳戶 ID。

```
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
}
```

使用混合存取模式共用 AWS Glue 資源

在另一個強制 Lake Formation AWS 帳戶 許可中與另一個 AWS 帳戶 或委託人共用資料，而不會中斷現有 Data Catalog 使用者的 IAM 型存取。

案例描述 - 生產者帳戶具有 Data Catalog 資料庫，該資料庫具有使用 Amazon S3 的 IAM 主體政策和 AWS Glue 動作控制的存取。資料庫的資料位置未向 Lake Formation 註冊。根據預設，IAMAllowedPrincipals群組具有資料庫及其所有資料表的Super許可。

在混合存取模式中授予跨帳戶 Lake Formation 許可

1. 製作者帳戶設定

   1. 使用具有 IAM 許可的角色登入 Lake Formation `lakeformation:PutDataLakeSettings` 主控台。
   2. 前往 Data Catalog 設定，然後選擇`Version 4`跨帳戶版本設定。

      如果您目前正在使用版本 1 或 2，請參閱更新至版本 3 更新跨帳戶資料共用版本設定的說明。

      從第 3 版升級至第 4 版時，不需要變更許可政策。
   3. 註冊您計劃在混合存取模式中共用的資料庫或資料表的 Amazon S3 位置。
   4. 在上述步驟中，確認您以混合存取模式註冊資料位置的資料庫和資料表上存在`IAMAllowedPrincipals`群組的`Super`許可。
   5. 將 Lake Formation 許可授予 AWS 組織、組織單位 (OUs)，或直接授予另一個帳戶中的 IAM 主體。
   6. 如果您要將許可直接授予 IAM 主體，請從消費者帳戶選擇加入主體，透過啟用選項讓 Lake Formation 許可立即生效，在混合存取模式中強制執行 Lake Formation 許可。

      如果您要將跨帳戶許可授予另一個 AWS 帳戶，當您選擇加入帳戶時，Lake Formation 許可只會對該帳戶的管理員強制執行。收件人帳戶資料湖管理員需要串聯許可，並選擇帳戶中的主體，以對處於混合存取模式的共用資源強制執行 Lake Formation 許可。

      如果您選擇 LF 標籤相符的資源選項來授予跨帳戶許可，則需要先完成授予許可步驟。您可以在 Lake Formation 主控台左側導覽列的許可下選擇混合存取模式，以選擇將主體和資源加入混合存取模式作為單獨的步驟。然後選擇新增以新增您要強制執行 Lake Formation 許可的資源和主體。

2. 消費者帳戶設定

   1. 以資料湖管理員身分登入 Lake Formation 主控台，網址為 https：//https://console.aws.amazon.com/lakeformation/。
   2. 前往 https：//https://console.aws.amazon.com/ram/home。 AWS RAM 主控台中的與我共用索引標籤會顯示與您的帳戶共用的資料庫和資料表。
   3. 在 Lake Formation 中建立共用資料庫和/或資料表的資源連結。
   4. 將資源連結和`Grant on target`許可（在原始共用資源上）的`Describe`許可授予您（消費者）帳戶中的 IAM 主體。

5. 將與您共用的資料庫或資料表的 Lake Formation 許可授予您帳戶中的委託人。選擇主體和資源，透過啟用讓 Lake Formation 許可立即生效的選項，在混合存取模式中強制執行 Lake Formation 許可。
6. 執行範例 Athena 查詢，以測試委託人的 Lake Formation 許可。使用 Amazon S3 和 AWS Glue 動作的 IAM 主體政策來測試 AWS Glue 使用者的現有存取權。

   （選用）移除您設定為使用 Lake Formation 許可之主體的資料存取 Amazon S3 儲存Amazon S3貯體政策和 的 AWS Glue IAM 主體政策，以及 Amazon S3 資料存取。

使用混合存取模式共用 Lake Formation 資源

允許外部帳戶中的新 Data Catalog 使用者使用 IAM 型政策存取 Data Catalog 資料庫和資料表，而不會中斷現有的 Lake Formation 跨帳戶共用許可。

案例描述 - 生產者帳戶具有 Lake Formation 受管資料庫和資料表，這些資料庫和資料表會在帳戶層級或 IAM 委託人層級與外部（消費者）帳戶共用。資料庫的資料位置已向 Lake Formation 註冊。`IAMAllowedPrincipals` 群組沒有資料庫及其資料表的`Super`許可。

透過以 IAM 為基礎的政策將跨帳戶存取權授予新的 Data Catalog 使用者，而不會中斷現有的 Lake Formation 許可

1. 製作者帳戶設定

   1. 使用 的角色登入 Lake Formation 主控台`lakeformation:PutDataLakeSettings`。
   2. 在資料目錄設定下，`Version 4`選擇跨帳戶版本設定。

      如果您目前正在使用版本 1 或 2，請參閱更新至版本 3 更新跨帳戶資料共用版本設定的說明。

      從第 3 版升級至第 4 版不需要任何許可政策變更。
   3. 列出您已授予資料庫和資料表主體的許可。如需詳細資訊，請參閱在 Lake Formation 中檢視資料庫和資料表許可。
   4. 選擇加入主體和資源，以授予現有的 Lake Formation 跨帳戶許可。

      > **Note**
      >
      > 在將資料位置註冊更新為混合存取模式以授予跨帳戶許可之前，您需要為每個帳戶授予至少一個跨帳戶資料共用。此步驟是更新附加至 AWS RAM 資源共享的 AWS RAM 受管許可的必要步驟。

2023 年 7 月，Lake Formation 已更新用於共用資料庫和資料表的 AWS RAM 受管許可：

- `arn:aws:ram::aws:permission/AWSRAMLFEnabledGlueAllTablesReadWriteForDatabase`（資料庫層級共享政策）
- `arn:aws:ram::aws:permission/AWSRAMLFEnabledGlueTableReadWrite`（資料表層級共享政策）

在 2023 年 7 月之前進行的跨帳戶許可授予沒有這些更新的 AWS RAM 許可。
如果您已將跨帳戶許可直接授予委託人，則需要個別將這些許可授予委託人。如果您略過此步驟，存取共用資源的委託人可能會收到非法的組合錯誤。

5. 前往 https：//https://console.aws.amazon.com/ram/home。
6. AWS RAM 主控台中的由我共用索引標籤會顯示您已與外部帳戶或委託人共用的資料庫和資料表名稱。

   確定連接至共用資源的許可具有正確的 ARN。
7. 驗證 AWS RAM 共享中的資源處於 `Associated` 狀態。如果狀態顯示為 `Associating`，請等待其進入 `Associated` 狀態。如果狀態變成 `Failed`，請停止並聯絡 Lake Formation 服務團隊。
8. 在左側導覽列的許可下選擇混合存取模式，然後選擇新增。
9. 新增主體和資源頁面會顯示資料庫和/或資料表，以及有權存取的主體。您可以新增或移除主體和資源，以進行必要的更新。
10. 針對您要變更為混合存取模式的資料庫和資料表，選擇具有 Lake Formation 許可的主體。選擇資料庫和資料表。
11. 選擇新增以選擇加入主體，以在混合存取模式中強制執行 Lake Formation 許可。
12. 將`Super`許可授予資料庫和所選資料表`IAMAllowedPrincipals`上的虛擬群組。
13. 將 Amazon S3 位置 Lake Formation 註冊編輯為混合存取模式。
14. 使用 Amazon S3 AWS Glue actions 的 IAM 許可政策，為外部（消費者）帳戶中 AWS Glue 的使用者授予許可。

2. 消費者帳戶設定

   1. 以資料湖管理員身分登入 Lake Formation 主控台，網址為 https：//https://console.aws.amazon.com/lakeformation/。
   2. 前往 https：//https://console.aws.amazon.com/ram/home。 AWS RAM 頁面中與我共用的資源索引標籤會顯示與您的帳戶共用的資料庫和資料表名稱。

針對 AWS RAM 共用，請確定連接的許可具有共用 AWS RAM 邀請的正確 ARN。檢查 AWS RAM 共用中的資源是否處於 `Associated` 狀態。如果狀態顯示為 `Associating`，請等待其進入 `Associated` 狀態。如果狀態變成 `Failed`，請停止並聯絡 Lake Formation 服務團隊。

3. 在 Lake Formation 中建立共用資料庫和/或資料表的資源連結。
4. 將資源連結和`Grant on target`許可（在原始共用資源上）的`Describe`許可授予您（消費者）帳戶中的 IAM 主體。
5. 接下來，在共用資料庫或資料表上為您帳戶中的主體設定 Lake Formation 許可。

   在左側導覽列的許可下，選擇混合存取模式。
6. 在混合存取模式頁面的下區段中選擇新增，以選擇加入委託人，以及從生產者帳戶與您共用的資料庫或資料表。
7. 使用 Amazon S3 AWS Glue actions 的 IAM 許可政策，為您帳戶中 AWS Glue 的使用者授予許可。
8. 使用 Athena 在資料表上執行個別的範例查詢，以測試使用者的 Lake Formation AWS Glue 許可

   （選用）針對處於混合存取模式的主體，清除 Amazon S3 的 IAM 許可政策。

從混合存取模式移除主體和資源

請依照下列步驟，從混合存取模式移除資料庫、資料表和主體。

Console

1. 登入 Lake Formation 主控台，網址為 https：//https://console.aws.amazon.com/lakeformation/。
2. 在許可下，選擇混合存取模式。
3. 在混合存取模式頁面上，選取資料庫或資料表名稱旁的核取方塊，然後選擇 `Remove`。
4. 警告訊息會提示您確認動作。選擇移除。

   Lake Formation 不會再強制執行這些資源的許可，而此資源的存取權將使用 IAM 和 AWS Glue 許可進行控制。如果使用者沒有適當的 IAM 許可，這可能會導致他們無法再存取此資源。

AWS CLI

下列範例示範如何從混合存取模式移除資源。

```
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>"
          }
    }
}
```

在混合存取模式中檢視主體和資源

請依照下列步驟，在混合存取模式中檢視資料庫、資料表和主體。

Console

1. 登入 Lake Formation 主控台，網址為 https：//https://console.aws.amazon.com/lakeformation/。
2. 在許可下，選擇混合存取模式。
3. 混合存取模式頁面顯示目前處於混合存取模式的資源和主體。

AWS CLI

下列範例顯示如何列出處於混合存取模式的所有選擇加入主體和資源。

```
aws lakeformation list-lake-formation-opt-ins
```

下列範例示範如何列出特定委託人資源對的選擇加入。

```
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>"
          }
    }
}
```

其他資源

在下列部落格文章中，我們會逐步解說如何在資料庫可透過 IAM 和 Amazon S3 許可存取其他使用者時，以混合存取模式為所選使用者加入 Lake Formation 許可的指示。我們將檢閱在帳戶內和兩個 AWS 帳戶之間設定混合存取模式的指示。

- 介紹 的混合存取模式 AWS Glue Data Catalog ，以使用 Lake Formation 和 IAM 和 Amazon S3 政策來保護存取。

在 中建立物件 AWS Glue Data Catalog

AWS Lake Formation 使用 AWS Glue Data Catalog (Data Catalog) 來儲存有關資料湖、資料來源、轉換和目標的中繼資料。中繼資料是資料集中的基礎資料的相關資料。每個 AWS 帳戶每個 AWS 區域都有一個資料目錄。

Data Catalog 中的中繼資料會組織在包含目錄、資料庫和資料表的三層資料階層中。它會將來自各種來源的資料組織到稱為目錄的邏輯容器中。每個目錄代表來自 Amazon Redshift 資料倉儲、 Amazon DynamoDB 資料庫和第三方資料來源的資料，例如 Snowflake、MySQL，以及透過聯合連接器整合的 30 多個外部資料來源。您也可以在資料目錄中建立新的目錄，將資料存放在 S3 資料表儲存貯體或 Redshift 受管儲存 (RMS) 中。

資料表會儲存基礎資料的相關資訊，包括結構描述資訊、分割區資訊和資料位置。資料庫是資料表的集合。Data Catalog 也包含資源連結，這些連結是外部帳戶中共用目錄、資料庫和資料表的連結，用於跨帳戶存取資料湖中的資料。

Data Catalog 是巢狀目錄物件，其中包含目錄、資料庫和資料表。它由 AWS 帳戶 ID 參考，並且是 帳戶和 中的預設目錄 AWS 區域。Data Catalog 使用三層階層 (catalog.database.table) 來組織資料表。

- Catalog – Data Catalog 三層中繼資料階層的最上層。您可以透過聯合在 Data Catalog 中新增多個目錄。
- 資料庫 – 中繼資料階層的第二層，包含資料表和檢視。在 Amazon Redshift 和 Trino 等許多資料系統中，資料庫也稱為結構描述。
- 資料表和檢視 – Data Catalog 的 3 層資料階層的第三層。

Amazon S3 中的所有 Iceberg 資料表都會存放在目錄 ID = AWS 帳戶 ID 的預設 Data Catalog 中。您可以在 中建立聯合目錄 AWS Glue Data Catalog ，透過聯合在 Amazon Redshift、Amazon S3 Table 儲存體或其他第三方資料來源中存放資料表的定義。

主題

- 建立目錄
- 建立資料庫
- 建立資料表
- 建置 AWS Glue Data Catalog 檢視

建立目錄

目錄代表 的三層中繼資料階層中最高層級或最高層級 AWS Glue Data Catalog。您可以使用多種方法來將資料帶入 Data Catalog 並建立多層級目錄。

如需從外部資料來源建立目錄的詳細資訊，請參閱 將您的資料帶入 AWS Glue Data Catalog。

若要使用 Lake Formation 主控台建立目錄，您必須以資料湖管理員或目錄建立者身分登入。目錄建立者是已獲授予 Lake Formation CREATE_CATALOG許可的主體。您可以在 Lake Formation 主控台的管理角色和任務頁面上查看目錄建立者清單。若要檢視此清單，您必須擁有 lakeformation:ListPermissions IAM 許可，並以資料湖管理員或目錄建立者身分登入，並在CREATE_CATALOG許可上使用 授予選項。

建立資料庫

Data Catalog 中的中繼資料表會存放在資料庫中。您可以視需要建立任意數量的資料庫，也可以對每個資料庫授予不同的 Lake Formation 許可。

資料庫可以有選用的位置屬性。此位置通常位於向 Lake Formation 註冊的 Amazon Simple Storage Service (Amazon S3) 位置。當您指定位置時，主體不需要資料位置許可，即可建立指向資料庫位置內位置的 Data Catalog 資料表。如需詳細資訊，請參閱Underlying data access control。

若要使用 Lake Formation 主控台建立資料庫，您必須以資料湖管理員或資料庫建立者身分登入。資料庫建立者是已獲授予 Lake Formation CREATE_DATABASE許可的主體。您可以在 Lake Formation 主控台的管理角色和任務頁面上查看資料庫建立者清單。若要檢視此清單，您必須擁有 `lakeformation:ListPermissions` IAM 許可，並以資料湖管理員或資料庫建立者身分登入，並具有CREATE_DATABASE許可上的授予選項。

若要建立資料庫

1. 在 https：//https://console.aws.amazon.com/lakeformation/ 開啟 AWS Lake Formation 主控台，並以資料湖管理員或資料庫建立者身分登入。
2. 在導覽窗格中的資料目錄下，選擇資料庫。
3. 選擇建立資料庫。
4. 在建立資料庫對話方塊中，輸入資料庫名稱、選用位置和選用描述。
5. 或者，選取僅針對此資料庫中的新資料表使用 IAM 存取控制。

   如需此選項的詳細資訊，請參閱the section called “變更資料湖的預設設定”。
6. 選擇建立資料庫。

建立資料表

AWS Lake Formation 中繼資料表包含資料湖中資料的相關資訊，包括結構描述資訊、分割區資訊和資料位置。這些資料表存放在 AWS Glue Data Catalog 中。您可以使用它們來存取資料湖中的基礎資料，並使用 Lake Formation 許可管理該資料。資料表存放在 Data Catalog 的資料庫中。

有多種方法可以建立 Data Catalog 資料表：

- 在 中執行爬蟲程式AWS Glue。請參閱《 AWS Glue 開發人員指南》中的定義爬蟲程式。
- 建立並執行工作流程。請參閱 the section called “使用工作流程匯入資料”。

- 使用 Lake Formation 主控台、AWS GlueAPI 或 AWS Command Line Interface () 手動建立資料表 AWS CLI。
- 使用 建立資料表 Amazon Athena。
- 建立外部帳戶中資料表的資源連結。請參閱 the section called “建立資源連結”。

建立 Apache Iceberg 資料表

AWS Lake Formation 支援建立 Apache Iceberg 資料表，在 中使用 Apache Parquet 資料格式 AWS Glue Data Catalog ，並在 Amazon S3 中使用資料。Data Catalog 中的資料表是中繼資料定義，代表資料存放區中的資料。根據預設，Lake Formation 會建立 Iceberg v2 資料表。有關 v1 和 v2 資料表之間的區別，請參閱 Apache Iceberg 文件中的格式版本變更。

Apache Iceberg 是開放式的資料表格式，專用於非常大型的分析資料集。Iceberg 可讓您輕鬆地變更結構描述，也稱為結構描述演變，這表示使用者可以從資料表新增、重新命名或移除資料欄，而不會中斷基礎資料。Iceberg 也支援資料版本控制，可讓使用者追蹤資料加班的變更。這可啟用時間歷程功能，允許使用者存取和查詢資料的歷史版本，並分析更新和刪除之間的資料變更。

您可以使用 Lake Formation 主控台或 AWS Glue API 中的 `CreateTable`操作，在 Data Catalog 中建立 Iceberg 資料表。如需詳細資訊，請參閱 CreateTable 動作 (Python：create_table)。

當您在 Data Catalog 中建立 Iceberg 資料表時，您必須在 Amazon S3 中指定資料表格式和中繼資料檔案路徑，才能執行讀取和寫入。

當您向 Amazon S3 資料位置註冊時，您可以使用 Lake Formation 使用精細存取控制許可來保護 Iceberg 資料表 AWS Lake Formation。對於 Amazon S3 中的來源資料和未向 Lake Formation 註冊的中繼資料，存取權取決於 Amazon S3 和 AWS Glue 動作的 IAM 許可政策。如需詳細資訊，請參閱管理 Lake Formation 許可。

> **Note**
>
> Data Catalog 不支援建立分割區和新增 Iceberg 資料表屬性。

主題

- 先決條件
- 建立 Iceberg 資料表

先決條件

若要在 Data Catalog 中建立 Iceberg 資料表，並設定 Lake Formation 資料存取許可，您需要完成下列要求：

1. 建立 Iceberg 資料表所需的許可，而沒有向 Lake Formation 註冊的資料。

   除了在 Data Catalog 中建立資料表所需的許可之外，資料表建立者還需要下列許可：

   - `s3:PutObject` 資源 arn：aws：s3：：：{bucketName} 上的
   - `s3:GetObject` 資源 arn：aws：s3：：：{bucketName} 上的
   - `s3:DeleteObject`資源 arn：aws：s3：：：{bucketName} 上的
2. 使用向 Lake Formation 註冊的資料建立 Iceberg 資料表所需的許可：

   若要使用 Lake Formation 來管理和保護資料湖中的資料，請使用 Lake Formation 註冊具有資料表資料的 Amazon S3 位置。這可讓 Lake Formation 將登入資料提供給 AWS 分析服務，例如 Athena、Redshift Spectrum 和 Amazon EMR 來存取資料。如需註冊 Amazon S3 位置的詳細資訊，請參閱 將 Amazon S3 位置新增至您的資料湖。

   讀取和寫入向 Lake Formation 註冊之基礎資料的委託人需要下列許可：

   - `lakeformation:GetDataAccess`
   - `DATA_LOCATION_ACCESS`

     在位置上具有資料位置許可的委託人也在所有子位置上具有位置許可。

     如需資料位置許可的詳細資訊，請參閱 基礎資料存取控制。

若要啟用壓縮，服務需要擔任具有更新 Data Catalog 中資料表許可的 IAM 角色。如需詳細資訊，請參閱資料表最佳化先決條件。

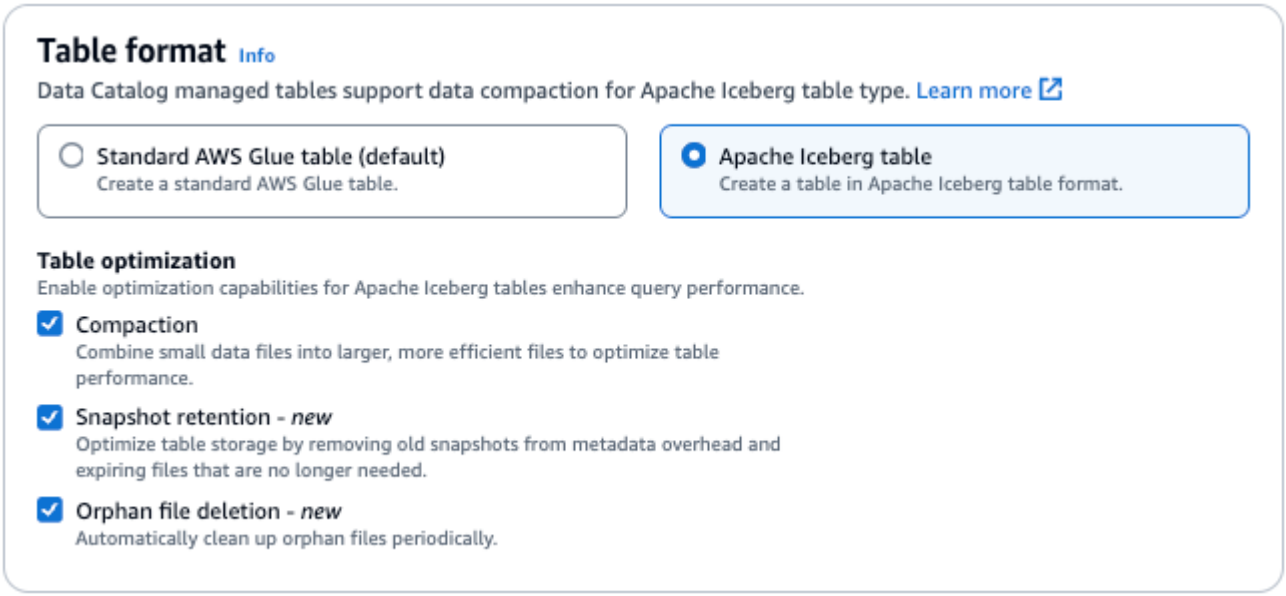
建立 Iceberg 資料表

您可以使用 Lake Formation 主控台或此頁面所記載 AWS Command Line Interface 的方式建立 Iceberg v1 和 v2 資料表。您也可以使用 AWS Glue 主控台或 建立 Iceberg 資料表 AWS Glue 編目程式。如需詳細資訊，請參閱《 AWS Glue 開發人員指南》中的資料目錄和爬蟲程式。

建立 Iceberg 資料表

Console

1. 登入 AWS Management Console，並在 https：//https://console.aws.amazon.com/lakeformation/ 開啟 Lake Formation 主控台。
2. 在資料型錄下，選擇資料表，然後使用建立資料表按鈕來指定下列屬性：
   - 資料表名稱：輸入資料表的名稱。如果您使用 Athena 存取資料表，請使用《Amazon Athena 使用者指南》中的這些命名提示。
   - 資料庫：選擇現有的資料庫或建立新的資料庫。
   - 描述：資料表的描述。您可以撰寫說明，來協助您了解資料表的內容。
   - 資料表格式：針對資料表格式，選擇 Apache Iceberg。



   - 資料表最佳化
     - 壓縮 – 合併和重寫資料檔案會移除過時的資料，並將分段的資料合併成更大、更有效率的檔案。
     - 快照保留 – 快照是 Iceberg 資料表的時間戳記版本。快照保留組態可讓客戶強制執行保留快照的時間長度，以及要保留的快照數量。設定快照保留最佳化工具可透過移除較舊、不必要的快照及其相關聯的基礎檔案，協助管理儲存體額外負荷。
     - 孤立檔案刪除 – 孤立檔案是 Iceberg 資料表中繼資料不再參考的檔案。這些檔案可能會隨著時間累積，特別是在資料表刪除或失敗的 ETL 任務等操作之後。啟用孤立檔案刪除 AWS Glue 可讓 定期識別和移除這些不必要的檔案，釋放儲存空間。

     如需詳細資訊，請參閱最佳化 Iceberg 資料表。

- IAM 角色：若要執行壓縮，服務會代表您擔任 IAM 角色。您可以使用下拉式選單選擇 IAM 角色。請確認角色具有啟用壓縮的必要權限。

  若要進一步了解所需的許可，請參閱資料表最佳化先決條件。

- 位置：指定 Amazon S3 中存放中繼資料表之資料夾的路徑。Iceberg 需要資料目錄中的中繼資料檔案和位置，才能執行讀取和寫入。

- 結構描述：選擇新增資料欄以新增資料欄和資料欄的資料類型。您可以選擇建立空白資料表，稍後再更新結構描述。Data Catalog 支援 Hive 資料類型。如需詳細資訊，請參閱 Hive 資料類型。

  Iceberg 可讓您在建立資料表之後發展結構描述和分割區。您可以使用 Athena 查詢來更新資料表結構描述，以及更新分割區的 Spark 查詢。

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/"
            }
        }'
```

最佳化處理 Iceberg 資料表

Lake Formation 支援多個資料表最佳化選項，以增強 AWS 分析引擎和 ETL 任務所使用的 Apache Iceberg 資料表管理和效能。這些最佳化工具提供高效率的儲存使用率、改善的查詢效能，以及有效的資料管理。Lake Formation 提供三種類型的資料表最佳化工具：

- 壓縮 – 資料壓縮壓縮小型資料檔案，以減少儲存用量並改善讀取效能。資料檔案會合併並重新寫入，以移除過時的資料，並將分段的資料合併成更大、更有效率的檔案。壓縮可設定為自動執行，或視需要手動觸發。
- 快照保留 – 快照是 Iceberg 資料表的時間戳記版本。快照保留組態可讓客戶強制執行保留快照的時間長度，以及要保留的快照數量。設定快照保留最佳化工具可透過移除較舊、不必要的快照及其相關聯的基礎檔案，協助管理儲存體額外負荷。
- 孤立檔案刪除 – 孤立檔案是 Iceberg 資料表中繼資料不再參考的檔案。這些檔案可能會隨著時間累積，特別是在資料表刪除或失敗的 ETL 任務等操作之後。啟用孤立檔案刪除 AWS Glue 可讓 定期識別和移除這些不必要的檔案，釋放儲存空間。

您可以使用 AWS Glue 主控台或 AWS Glue API 操作，啟用或停用 Data Catalog 中個別 Iceberg 資料表的壓縮 AWS CLI、快照保留和孤立檔案刪除最佳化工具。

如需詳細資訊，請參閱《 AWS Glue 開發人員指南》中的最佳化 Iceberg 資料表。

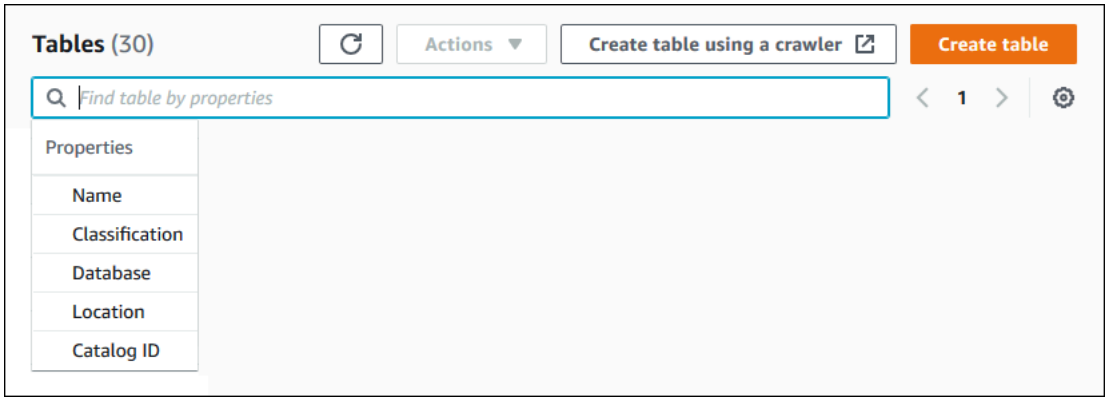
搜尋資料表

您可以使用 AWS Lake Formation 主控台，依名稱、位置、包含資料庫等來搜尋 Data Catalog 資料表。搜尋結果只會顯示您擁有 Lake Formation 許可的資料表。

搜尋資料表 （主控台）

1. 登入 AWS Management Console 並開啟 Lake Formation 主控台，網址為 https：//https://console.aws.amazon.com/lakeformation/。
2. 在導覽窗格中，選擇 Tables (資料表)。
3. 將游標放在頁面頂端的搜尋欄位中。欄位具有依屬性尋找資料表的預留位置文字。

   屬性功能表隨即出現，顯示要搜尋的各種資料表屬性。



4. 執行以下任意一項：

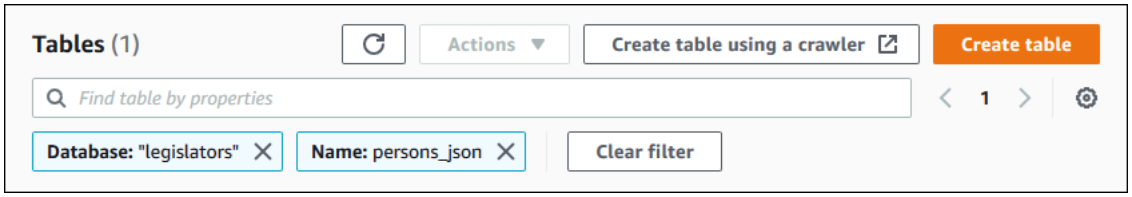
- 依包含資料庫的 搜尋。

  1. 從屬性功能表中選擇資料庫，然後從出現的資料庫功能表中選擇資料庫，或輸入資料庫名稱，然後按 Enter 鍵。

     列出您在資料庫中擁有許可的資料表。

  2. （選用）若要將清單縮小為資料庫中的單一資料表，請再次在搜尋欄位中放置游標，從屬性功能表中選擇名稱，然後從出現的資料表功能表中選擇資料表名稱，或輸入資料表名稱，然後按 Enter 鍵。

     系統會列出單一資料表，資料庫名稱和資料表名稱都會在搜尋欄位下顯示為圖磚。

     若要調整篩選條件，請關閉其中一個圖磚，或選擇清除篩選條件。

- 依其他屬性搜尋。

  1. 從屬性功能表中選擇搜尋屬性。

     若要依 AWS 帳戶 ID 搜尋，請從屬性功能表中選擇目錄 ID，輸入有效的 AWS 帳戶 ID（例如 111122223333)，然後按 Enter。

     若要依位置搜尋，請從屬性功能表中選擇位置，然後從出現的位置功能表中選取位置。系統會傳回所選位置（例如 Amazon S3) 根位置中的所有資料表。

使用 搜尋資料表 AWS CLI

- 下列範例示範如何執行部分 serach。`--search-text` 參數可讓您搜尋在其中繼資料中包含指定文字的資料表。在這種情況下，它會傳回名稱、描述或其他中繼資料欄位中具有「客戶」的所有資料表。

```
aws glue search-tables
      --search-text "customer"
      --region AWS ##
      --max-results 10
      --sort-criteria "FieldName=Name,Sort=ASC"
```

跨 AWS 帳戶共用 Data Catalog 資料表和資料庫

您可以透過將資源的 Lake Formation 許可授予外部 AWS 帳戶，與外部帳戶共用 Data Catalog 資源（資料庫和資料表）。然後，使用者可以執行查詢和任務，以跨多個帳戶聯結和查詢資料表。在某些限制下，當您與其他帳戶共用 Data Catalog 資源時，該帳戶中的主體可以在該資源上操作，就像資源在其 Data Catalog 中一樣。

您不與外部 AWS 帳戶中的特定主體共用資源，而是與 AWS 帳戶或組織共用資源。當您與 AWS 組織共用資源時，您要與該組織所有層級的所有帳戶共用資源。然後，每個外部帳戶中的資料湖管理員必須將共用資源的許可授予其帳戶中的主體。

如需詳細資訊，請參閱Lake Formation 中的跨帳戶資料共用及授予 Data Catalog 資源的許可。

> 另請參閱:
>
> - 存取和檢視共用的 Data Catalog 資料表和資料庫
> - 先決條件

建置 AWS Glue Data Catalog 檢視

在 中 AWS Glue Data Catalog，檢視是虛擬資料表，其中內容是由參考一或多個資料表的 SQL 查詢定義。您可以使用 Amazon Athena、Amazon Redshift 或 的 SQL 編輯器，或使用 EMR Serverless 或 5.0 AWS Glue 版的 Apache Spark，建立最多參考 10 個資料表的資料目錄檢視。檢視的基礎參考資料表可以屬於相同 AWS 帳戶 Data Catalog 中的相同資料庫或不同資料庫。

您可以參考開放 AWS Glue 資料表格式 (OTF) 的標準資料表和資料表，例如 Apache Hudi、Linux Foundation Delta Lake 和 Apache Iceberg，並將基礎資料存放在向 註冊的 Amazon S3 位置 AWS Lake Formation。此外，您可以從與 Lake Formation 共用的 Amazon Redshift 資料共用，從聯合資料表建立檢視。

區分 Data Catalog 檢視與其他檢視類型

Data Catalog 檢視與 Apache Hive、Apache Spark 和 Amazon Athena 檢視不同。Data Catalog 檢視是 的原生功能 AWS Glue Data Catalog，也是多向度定義者建立的檢視。您可以使用其中一個支援的分析服務建立 Data Catalog 檢視，例如 Athena 或 Amazon Redshift Spectrum，並使用其他支援的分析服務存取相同的檢視。另一方面，Apache Hive、Apache Spark 和 Athena 檢視會在 Athena 和 Amazon Redshift 等每個分析服務中獨立建立，並且只能在該服務內可見和存取。

什麼是定義器檢視？

定義者檢視是一種 SQL 檢視，根據建立它的委託人許可運作。定義者角色具有存取參考資料表的必要許可，並執行定義檢視的 SQL 陳述式。定義者會建立檢視，並透過 AWS Lake Formation精細存取控制與其他使用者共用檢視。

當使用者查詢定義者檢視時，查詢引擎會使用定義者角色的許可來存取基礎參考資料表。此方法可讓使用者與檢視互動，而無需直接存取來源資料表、增強安全性並簡化資料存取管理。

若要設定定義者檢視，定義者必須是在其資料目錄中託管檢視之相同 AWS 帳戶中的 IAM 角色。如需定義者角色所需許可的詳細資訊，請參閱 建立檢視的先決條件。

多方向檢視的架構

Data Catalog 支援使用多種結構化查詢語言 (SQL) 方言建立檢視。SQL 是一種用於在關聯式資料庫中存放和處理資訊的語言，每個 AWS 分析引擎都會使用自己的 SQL 或 SQL 方言變化。

您可以使用其中一個支援的分析查詢引擎，在一個 SQL 方言中建立 Data Catalog 檢視。然後，您可以使用任何其他支援的分析引擎中不同 SQL 方言中的 `ALTER VIEW`陳述式來更新檢視。不過，每個方言都必須參考一組相同的資料表、資料欄和資料類型。

您可以使用 `GetTable` API AWS CLI 和 AWS 主控台存取可用於檢視的多個方言。因此，資料目錄檢視是可見的，可用於跨不同支援的分析引擎進行查詢。

透過定義您可以從多個引擎查詢的常見檢視結構描述和中繼資料物件，Data Catalog 檢視可讓您跨資料湖使用統一檢視。

如需如何解決每個方言結構描述的詳細資訊，請參閱 API 參考的連結。如需不同類型相符規則的詳細資訊，請參閱 API 文件中的相關區段連結。

與 Lake Formation 許可整合

您可以使用 AWS Lake Formation 集中管理使用者 AWS Glue Data Catalog 檢視的許可。您可以使用具名資源方法或 LF 標籤，在 Data Catalog 檢視上授予精細許可，並跨 AWS 帳戶、 AWS 組織和組織單位共用這些許可。您也可以 AWS 區域 使用資源連結跨 共用和存取 Data Catalog 檢視。這可讓使用者提供資料存取，而無需複製資料來源，並共用基礎資料表。

Data Catalog 檢視的 `CREATE VIEW` DDL 陳述式可以開放資料表格式 (OTF) 參考標準 AWS Glue 資料表和資料表，例如 Hudi、Delta Lake 和 Iceberg，其中包含存放在向 Lake Formation 註冊的 Amazon S3 位置的基礎資料，以及與 Lake Formation 共用的 Amazon Redshift 資料共用的聯合資料表。只要用於查詢檢視的引擎支援該格式，資料表可以是任何檔案格式。您也可以參考執行引擎的內建函數，但可能不允許其他引擎特定資源。如需詳細資訊，請參閱Data Catalog 檢視的考量和限制。

使用案例

以下是 Data Catalog 檢視的重要使用案例：

- 建立和管理單一檢視結構描述的許可。這可協助您避免在多個引擎中建立的重複檢視上具有不一致許可的風險。
- 在參考多個資料表的檢視中將許可授予使用者，而不直接在基礎參考資料表上授予許可。
- 透過在檢視上套用 LF 標籤，並將以 LF 標籤為基礎的許可授予使用者，在資料表上使用 LF 標籤（其中 LF 標籤只會層疊至資料欄層級）來達成資料列層級篩選。

檢視支援的 AWS 分析服務

下列 AWS 分析服務支援建立 Data Catalog 檢視：

- Amazon Redshift
- Amazon Athena 第 3 版
- EMR Serverless 上的 Apache Spark
- 5.0 AWS Glue 版上的 Apache Spark

其他資源

您可以在本指南中進一步了解 Data Catalog，以及使用下列資源：

以下影片示範如何從 Athena 和 Amazon Redshift 建立檢視和查詢檢視。

主題

- 建立檢視的先決條件
- 使用 DDL 陳述式建立 Data Catalog 檢視
- 使用 AWS Glue APIs建立 Data Catalog 檢視
- 授予 Data Catalog 檢視的許可

建立檢視的先決條件

- 若要在 Data Catalog 中建立檢視，您必須向 Lake Formation 註冊參考資料表的基礎 Amazon S3 資料位置。如需向 Lake Formation 註冊資料的詳細資訊，請參閱 將 Amazon S3 位置新增至您的資料湖。
- 只有 IAM 角色可以建立 Data Catalog 檢視。其他 IAM 身分無法建立 Data Catalog 檢視。
- 定義檢視的 IAM 角色必須具有下列許可：
  - Lake Formation SELECT許可，其中包含所有參考資料表上的 Grantable選項。
  - Lake Formation 和 AWS Glue 服務擔任角色的信任政策。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "DataCatalogViewDefinerAssumeRole1",
            "Effect": "Allow",
            "Principal": {
               "Service": [
                    "glue.amazonaws.com",
                    "lakeformation.amazonaws.com"
                ]
            },
            "Action": "sts:AssumeRole"
        }
    ]
}
```

  - AWS Glue 和 Lake Formation 的 iam：PassRole 許可。

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

```
    "Statement": [
        {
            "Sid": "DataCatalogViewDefinerPassRole1",
            "Action": [
                "iam:PassRole"
            ],
            "Effect": "Allow",
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "iam:PassedToService": [
                        "glue.amazonaws.com",
                        "lakeformation.amazonaws.com"
                      ]
                }
            }
        }
    ]
}
```

- AWS Glue 和 Lake Formation 許可。

```
{
    "Version": "2012-10-17",
                 "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "Glue:GetDatabase",
                "Glue:GetDatabases",
                "Glue:CreateTable",
                "Glue:GetTable",
                "Glue:GetTables",
                "Glue:BatchGetPartition",
                "Glue:GetPartitions",
                "Glue:GetPartition",
                "Glue:GetTableVersion",
                "Glue:GetTableVersions",
    "Glue:PassConnection",
                "lakeFormation:GetDataAccess"
            ],
            "Resource": "*"
        }
    ]
```

```
}
```

- 您無法在已將 Super或 ALL許可授予IAMAllowedPrincipals群組的資料庫中建立檢視。您可以在資料庫上撤銷 IAMAllowedPrincipals群組的Super許可，請參閱 步驟 4：將資料存放區切換至 Lake Formation 許可模型，或使用在新建立資料表的預設許可下取消勾選，在此資料庫方塊中僅針對新資料表使用 IAM 存取控制來建立新的資料庫。

使用 DDL 陳述式建立 Data Catalog 檢視

您可以使用 Athena、Amazon Redshift 和 AWS Glue APIs/ 的 SQL 編輯器建立 AWS Glue Data Catalog 檢視AWS CLI。

若要使用 SQL 編輯器建立 Data Catalog 檢視，請選擇 Athena 或 Redshift Spectrum，然後使用CREATE VIEW資料定義語言 (DDL) 陳述式建立檢視。在第一個引擎的方言中建立檢視後，您可以使用第二個引擎的 ALTER VIEW DDL 陳述式來新增其他方言。

定義檢視時，請務必考量下列事項：

- 定義多方檢視 – 當您定義具有多個方言的檢視時，不同方言的結構描述必須相符。每個 SQL 方言都有略有不同的語法規格。定義 Data Catalog 檢視的查詢語法應該在所有方言中解析為完全相同的資料欄清單，包括類型和名稱。此資訊會存放在 檢視StorageDescriptor的 中。方言也必須參考來自 Data Catalog 的相同基礎資料表物件。

  若要使用 DDL 將另一個方言新增至檢視，您可以使用 ALTER VIEW陳述式。如果ALTER VIEW陳述式嘗試更新檢視定義，例如修改檢視的儲存描述項或基礎資料表，則陳述式會錯誤地顯示「輸入和現有儲存描述項不相符」。您可以使用 SQL 轉換操作來確保檢視欄類型相符。
- 更新檢視 – 若要更新檢視，您可以使用 UpdateTable API。如果您更新檢視時沒有相符的儲存描述項或參考資料表，您可以提供 FORCE旗標 （如需語法，請參閱引擎 SQL 文件）。強制更新後，檢視將採用強制資料表StorageDescriptor和參考資料表。任何進一步的 ALTER VIEW DDL 都應符合修改後的值。已更新為具有不相容方言的檢視將處於「過時」狀態。檢視狀態會顯示在 Lake Formation 主控台和使用 GetTable操作。
- 參考 varchar 資料欄類型做為字串 – 無法將 Redshift Spectrum 的 varchar 資料欄類型轉換為字串。如果在 Redshift Spectrum 中使用 varchar 資料欄類型建立檢視，且後續方言嘗試將該欄位參考為字串，則 Data Catalog 會將它視為字串，而不需要FORCE旗標。
- 複雜類型欄位的處理 – Amazon Redshift 會將所有複雜類型視為 SUPER 類型，而 Athena 會指定複雜類型。如果檢視具有SUPER類型欄位，且另一個引擎參考該資料欄作為特定複雜類型，例如 struct (<street_address:struct<street_number:int, street_name:string,

`street_type:string>>)`，則 Data Catalog 會假設該欄位為特定複雜類型，並在儲存體描述項中使用該欄位，而不需要`Force`旗標。

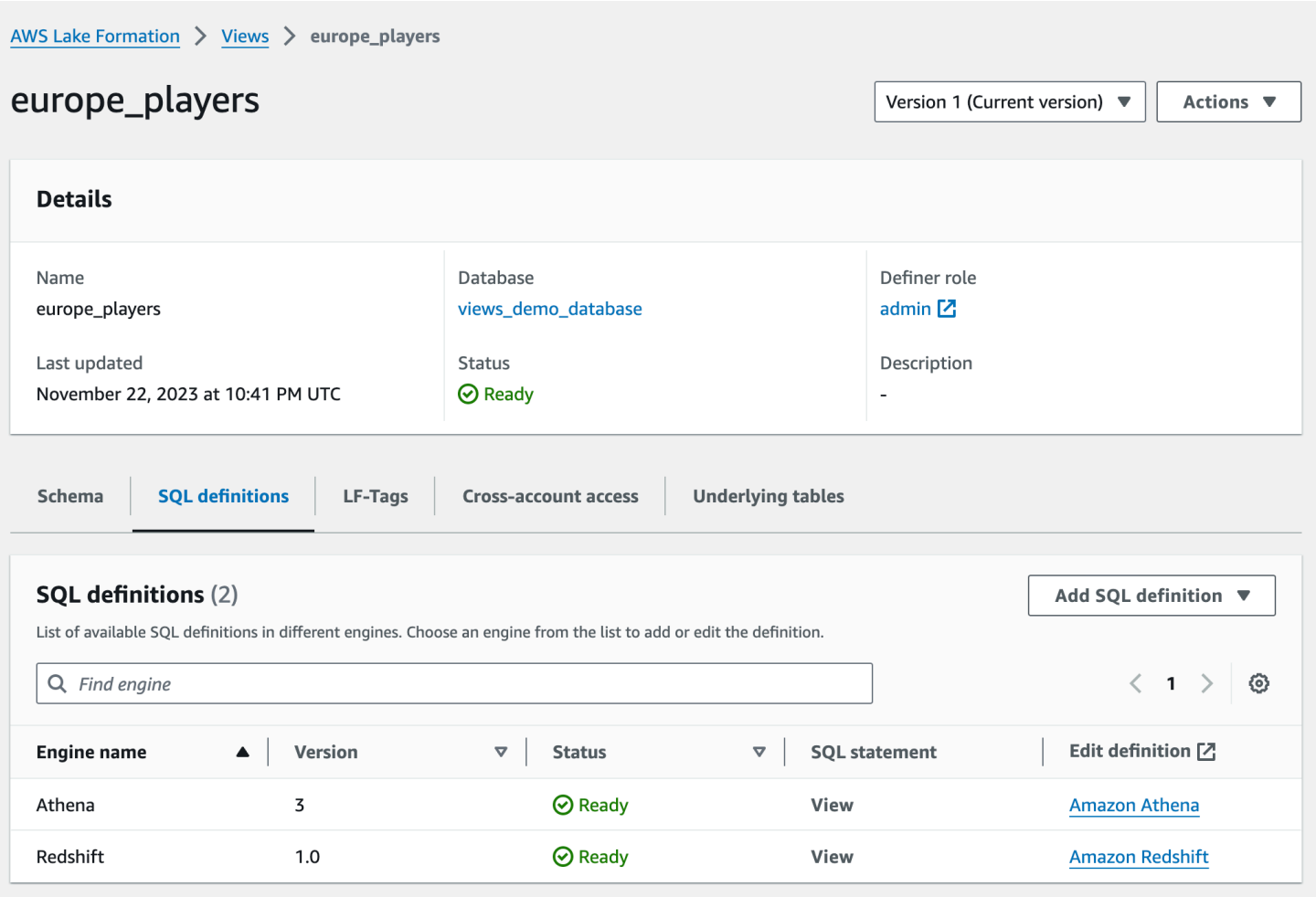
如需建立和管理 Data Catalog 檢視之語法的詳細資訊，請參閱：

- 《Amazon Athena 使用者指南》中的使用 AWS Glue Data Catalog 檢視。
- 《Amazon Athena 使用者指南》中的 Glue Data Catalog 檢視查詢語法。
- 《Amazon Redshift 資料庫開發人員指南》中的在 中建立檢視 AWS Glue Data Catalog。

若要取得有關與 Data Catalog 中的視觀表相關之 SQL 命令的更多資訊，請參閱CREATE EXTERNAL VIEW、ALTER EXTERNAL VIEW 和 DROP EXTERNAL VIEW。

建立 Data Catalog 檢視後，您可以在 Lake Formation 主控台中取得檢視的詳細資訊。

1. 在 Lake Formation 主控台中選擇資料目錄下的檢視。
2. 可用檢視的清單會顯示在檢視頁面上。
3. 從清單中選擇檢視，詳細資訊頁面會顯示檢視的屬性。



結構描述

選擇資料Column列，然後選取編輯 LF 標籤以更新標籤值或指派新的 LF 標籤。

SQL 定義

您可以查看可用的 SQL 定義清單。選取新增 SQL 定義，然後選擇查詢引擎以新增 SQL 定義。在資料Edit definition欄下選擇查詢引擎 (Athena 或 Amazon Redshift) 以更新 SQL 定義。

LF 標籤

選擇編輯 LF 標籤來編輯標籤的值或指派新標籤。您可以使用 LF 標籤來授予檢視的許可。

跨帳戶存取權

您可以查看已與您共用 Data Catalog 檢視的組織 AWS 帳戶和組織單位 (OUs) 清單。

基礎資料表

用於建立檢視的 SQL 定義中參考的基礎資料表會顯示在此索引標籤下方。

使用 AWS Glue APIs建立 Data Catalog 檢視

您可以使用 AWS Glue CreateTable 和 UpdateTable APIs Data Catalog 中建立和更新檢視。CreateTable 和 UpdateTable操作具有 的新TableInput結構ViewDefinition，而 SearchTables、GetTableGetTables、GetTableVersion、、 GetTableVersions操作在其檢視的輸出語法ViewDefinition中提供 。此外，GetTableAPI 輸出中有新的Status欄位。

有兩個新 AWS Glue 連線可用於驗證每個支援的查詢引擎 Amazon Athena 和 Amazon Redshift 的 SQL 方言。

與檢視搭配使用時， CreateTable和 UpdateTable APIs是非同步的。當使用多個 SQL 方言呼叫這些 APIs 時，系統會使用每個引擎驗證呼叫，以判斷該方言是否可以在該引擎上執行，以及每個方言的檢視結果結構描述是否相符。 AWS Glue 服務使用這些連線對分析引擎進行內部呼叫。這些呼叫會模擬引擎如何驗證在引擎上執行的 CREATE VIEW或 ALTER VIEW SQL DDL。

如果提供的 SQL 有效，且檢視方言之間的結構描述相符，則 AWS Glue API 會以原子方式遞交結果。Atomicity 允許建立或修改具有多個方言的檢視，而不會停機。

主題

- 建立 AWS Glue 連線以驗證狀態
- 驗證檢視產生狀態
- 非同步狀態和操作
- 檢視非同步操作期間的建立失敗案例

建立 AWS Glue 連線以驗證狀態

若要使用 或 UpdateTable操作建立CreateTable或更新 AWS Glue Data Catalog 檢視，您必須建立新的連線類型 AWS Glue 以進行驗證，並將其提供給支援的分析引擎。搭配 Athena 或 Amazon Redshift 使用 Data Catalog 檢視需要這些連線。您只能使用 AWS CLI、 AWS SDKs或 AWS Glue APIs建立這些連線。您無法使用 AWS Management Console 建立 AWS Glue 連線。

> **Note**
>
> 如果檢視定義者角色和呼叫 CreateTable或 UpdateTable 的角色不同，則兩者都需要其 IAM 政策陳述式中的glue:PassConnection許可。

如需詳細資訊，請參閱 create-connection AWS CLI 文件。

AWS CLI 用於建立連線的 命令

以下是用於建立連線的 AWS CLI 命令：

```
aws glue create-connection --region us-east-1
--endpoint-url https://glue.us-east-1.amazonaws.com
--cli-input-json file:///root/path/to/create-connection.json
```

AWS CLI 輸入 JSON

對於 Amazon Redshift：

```
{
    "CatalogId": "123456789012",
    "ConnectionInput": {
        "ConnectionType": "VIEW_VALIDATION_REDSHIFT",
        "Name": "views-preview-cluster-connection-2",
        "Description": "My first Amazon Redshift validation connection",
        "ConnectionProperties": {
            "DATABASE": "dev",
            "CLUSTER_IDENTIFIER": "glue-data-catalog-views-preview-cluster"
        }
    }
}
```

對於 Amazon Athena：

```
{
    "CatalogId": "123456789012",
    "ConnectionInput": {
        "ConnectionType": "VIEW_VALIDATION_ATHENA",
        "Name": "views-preview-cluster-connection-3",
        "Description": "My first Amazon Athena validation connection",
        "ConnectionProperties": {
            "WORKGROUP_NAME": "workgroup-name"
        }
    }
}
```

驗證檢視產生狀態

當您執行 `CreateTable`或 `UpdateTable`操作時，`GetTable`API 輸出的 `Status` 欄位會顯示檢視建立狀態的詳細資訊。對於資料表尚不存在的`create`請求， 會在非同步程序期間 AWS Glue 建立空資

料表。呼叫 時GetTable，您可以傳遞選用的布林值旗標 IncludeStatusDetails，其中顯示請求的診斷資訊。如果失敗，此旗標會顯示錯誤訊息，其中包含每個方言的個別狀態。

檢視建立、讀取、更新和刪除 (CRUD) 操作期間的錯誤可能發生在 AWS Glue/Lake Formation 服務的處理期間，或在 Amazon Redshift 或 Athena 的檢視 SQL 驗證期間。在引擎驗證期間發生錯誤時，AWS Glue 服務會提供引擎傳回的錯誤訊息。

狀態欄位

以下是狀態欄位：

- 狀態：一般狀態，與不同類型的任務無關：
  - QUEUED
  - IN_PROGRESS
  - 成功
  - 失敗
- 動作 – 指出資料表上呼叫的動作，目前只有 CREATE或 UPDATE操作可用。

  使用檢視時，區分 UPDATE和 CREATE操作非常重要。操作類型會決定您應該如何繼續查詢資料表。

  UPDATE 操作表示資料表已存在於 Data Catalog 中。在這種情況下，您可以繼續查詢先前建立的資料表，而不會發生任何問題。另一方面， CREATE 操作表示先前從未成功建立資料表。如果資料表標示為 CREATE，嘗試查詢將會失敗，因為資料表尚未存在於系統中。因此，在嘗試查詢資料表之前，請務必識別操作類型 (UPDATE 或 CREATE)。
- RequestedBy – 請求非同步變更之使用者的 ARN。
- UpdatedBy – 上次手動變更非同步變更程序的使用者 ARN，例如請求取消或修改。
- 錯誤 – 此欄位只會在狀態為 FAILED 時出現。這是父層級例外狀況訊息。每個方言可能會有不同的錯誤。
  - ErrorCode – 例外狀況的類型。
  - ErrorMessage – 例外狀況的簡短描述。
- RequestTime – ISO 8601 格式的日期字串，指出啟動變更的時間。
- UpdateTime – ISO 8601 格式的日期字串，指出上次更新狀態的時間。

非同步狀態和操作

當您執行glue:CreateTable請求時，會開始非同步建立 Data Catalog 檢視。在下列各節中，本文件說明glue:GetTable回應中可用的 AWS Glue 檢視Status的 。為了簡潔起見，本節省略完整回應。

```
{
    "Table": {
        ...
        "Status": {
            ...
            "Action": "CREATE",
            "State": "QUEUED",
        }
    }
}
```

上述兩個屬性都代表重要的診斷資訊，指出非同步操作的狀態，以及可在此檢視上執行的動作。以下是這些屬性可以採用的可能值。

1. `Status.Action`
   a. CREATE
   b. UPDATE
2. `Status.State`
   a. QUEUED
   b. IN_PROGRESS
   c. 成功
   d. 失敗

也請務必注意，Data Catalog 檢視上的某些更新不需要非同步操作。例如，可能想要更新資料表的 `Description` 屬性。由於這不需要任何非同步操作，因此產生的資料表中繼資料將沒有任何 `Status`，且屬性將為 `NULL`。

```
{
    "Table": {
        ...,
        "Description": "I changed this attribute!"
    }
}
```

接下來，本主題會探索上述狀態資訊如何影響可在 AWS Glue 檢視上執行的操作。

glue：CreateTable

相較於任何 Glue 資料表的 glue:CreateTable函數，此 API 沒有任何變更。 CreateTable 可能針對尚未存在的任何資料表名稱呼叫 。

glue：UpdateTable

此操作無法在具有下列狀態資訊的 AWS Glue 檢視上執行：

1. 動作 == CREATE 和狀態 == QUEUED
2. 動作 == CREATE 和狀態 == IN_PROGRESS
3. 動作 == CREATE 和狀態 == 失敗
4. 動作 == UPDATE 和狀態 == QUEUED
5. 動作 == UPDATE 和狀態 == IN_PROGRESS

總而言之，您只能在符合下列要求時更新 Data Catalog 檢視。

1. 第一次成功建立。
   a. 動作 == CREATE 和狀態 == 成功
2. 其在非同步更新操作後已達到終端狀態。
   a. 動作 == UPDATE 和狀態 == 成功
   b. 動作 == UPDATE 和狀態 == 失敗
3. 由於同步更新，它具有NULL狀態屬性。

glue：DeleteTable

相較於任何 AWS Glue 資料表的 glue:DeleteTable 函數，此操作沒有任何變更。無論其狀態為何，您都可以刪除 Data Catalog 檢視。

glue：GetTable

相較於任何 AWS Glue 資料表的 glue:GetTable 函數，此操作沒有任何變更。不過，在第一次成功建立 Data Catalog 檢視之前，您無法從分析引擎查詢 Data Catalog 檢視。 Action == CREATE and State == SUCCESS第一次成功建立 Data Catalog 檢視後，無論其狀態為何，您都可以查詢檢視。

> **Note**
>
> 本節中的所有資訊適用於所有資料表讀取 APIsGetTables，例如 GetTable、 和 SearchTables。

檢視非同步操作期間的建立失敗案例

下列範例代表可能因CreateTable或UpdateTable檢視 API 呼叫而產生的錯誤類型。它們並不詳盡，因為 SQL 查詢失敗的錯誤表面相當大。

案例 1：Amazon Redshift 查詢失敗

驗證期間，在資料目錄中找不到針對 Amazon Redshift 提供的查詢，其中包含拼寫錯誤的資料表名稱。產生的錯誤會顯示在檢視GetTable回應的 Status 欄位中。

GetTable 請求：

```
{
    "CatalogId": "123456789012",
    "DatabaseName": "async-view-test-db",
    "TableInput": {
        "Name": "view-athena-redshift-72",
        "Description": "This is an atomic operation",
        "StorageDescriptor": {
            "Columns": [
                { "Name": "col1", "Type": "int" },
                { "Name": "col2", "Type": "string" },
                { "Name": "col3", "Type": "double" }
            ]
        },
        "ViewDefinition": {
            "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
            "SubObjects": [ "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1" ],
            "Representations": [
                {
                    "Dialect": "ATHENA",
                    "DialectVersion": "3",
                    "ViewOriginalText": "SELECT * FROM \"gdc-view-playground-db\".
\"table_1\"",
                    "ValidationConnection": "athena-connection"
```

```
                },
                {
                    "Dialect": "REDSHIFT",
                    "DialectVersion": "1.0",
                    "ViewOriginalText": "SELECT * FROM \"gdc-view-playground-external-
schema\".\"table__1\";",
                    "ValidationConnection": "redshift-connection"
                }
            ]
        }
    }
}
```

GetTable 回應：

```
IncludeStatusDetails = FALSE
{
    "Table": {
        "Name": "view-athena-redshift-72",
        "DatabaseName": "async-view-test-db",
        "Description": "",
        "CreateTime": "2024-07-11T11:39:19-07:00",
        "UpdateTime": "2024-07-11T11:39:19-07:00",
        "Retention": 0,
        "ViewOriginalText": "",
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "IsMultiDialectView": false,
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:39:19-07:00",
            "UpdateTime": "2024-07-11T11:40:06-07:00",
            "Action": "CREATE",
            "State": "FAILED"
        }
    }
```

```
}

IncludeStatusDetails = TRUE
{
    "Table": {
        "Name": "view-athena-redshift-72",
        "DatabaseName": "async-view-test-db",
        "Description": "",
        "CreateTime": "2024-07-11T11:39:19-07:00",
        "UpdateTime": "2024-07-11T11:39:19-07:00",
        "Retention": 0,
        "ViewOriginalText": "",
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "IsMultiDialectView": false,
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:39:19-07:00",
            "UpdateTime": "2024-07-11T11:40:06-07:00",
            "Action": "CREATE",
            "State": "FAILED",
            "Error": {
                "ErrorCode": "QueryExecutionException",
                "ErrorMessage": "Error received during view SQL validation
 using a connection: [Connection Name: redshift-connection | Query Execution
 Id: ddb711d3-2415-4aa9-b251-6a76ab4f41b1 | Timestamp: Thu Jul 11 18:39:37 UTC
 2024]: Redshift returned error for the statement: ERROR: AwsClientException:
 EntityNotFoundException from glue - Entity Not Found"
            },
            "Details": {
                "RequestedChange": {
                    "Name": "view-athena-redshift-72",
                    "DatabaseName": "async-view-test-db",
                    "Description": "This is an atomic operation",
                    "Retention": 0,
                    "StorageDescriptor": {
                        "Columns": [
```

```
                            {
                                "Name": "col1",
                                "Type": "int"
                            },
                            {
                                "Name": "col2",
                                "Type": "string"
                            },
                            {
                                "Name": "col3",
                                "Type": "double"
                            }
                        ],
                        "Compressed": false,
                        "NumberOfBuckets": 0,
                        "SortColumns": [],
                        "StoredAsSubDirectories": false
                    },
                    "TableType": "VIRTUAL_VIEW",
                    "IsRegisteredWithLakeFormation": false,
                    "CatalogId": "123456789012",
                    "IsRowFilteringEnabled": false,
                    "VersionId": "-1",
                    "DatabaseId": "<databaseID>",
                    "ViewDefinition": {
                        "IsProtected": true,
                        "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
                        "SubObjects": [
                            "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1"
                        ],
                        "Representations": [
                            {
                                "Dialect": "ATHENA",
                                "DialectVersion": "3",
                                "ViewOriginalText": "SELECT * FROM \"gdc-view-
playground-db\".\"table_1\"",
                                "IsStale": false
                            },
                            {
                                "Dialect": "REDSHIFT",
                                "DialectVersion": "1.0",
                                "ViewOriginalText": "SELECT * FROM \"gdc-view-
playground-external-schema\".\"table__1\";",
```

```
                                "IsStale": false
                            }
                        ]
                    },
                    "IsMultiDialectView": true
                },
                "ViewValidations": [
                    {
                        "Dialect": "ATHENA",
                        "DialectVersion": "3",
                        "ViewValidationText": "SELECT * FROM \"gdc-view-playground-db
\".\"table_1\"",
                        "UpdateTime": "2024-07-11T11:40:06-07:00",
                        "State": "SUCCESS"
                    },
                    {
                        "Dialect": "REDSHIFT",
                        "DialectVersion": "1.0",
                        "ViewValidationText": "SELECT * FROM \"gdc-view-playground-
external-schema\".\"table__1\";",
                        "UpdateTime": "2024-07-11T11:39:37-07:00",
                        "State": "FAILED",
                        "Error": {
                            "ErrorCode": "QueryExecutionException",
                            "ErrorMessage": "Error received during view SQL validation
 using a connection: [Connection Name: redshift-connection | Query Execution Id:
 ddb711d3-2415-4aa9-b251-6a76ab4f41b1 | Timestamp: Thu
 Jul 11 18:39:37 UTC 2024]: Redshift returned error for the statement: ERROR:
 AwsClientException: EntityNotFoundException from glue - Entity Not Found"
                        }
                    }
                ]
            }
        }
    }
}
```

案例 2：無效的 Amazon Redshift 連線

下列範例中的 Amazon Redshift 連線格式錯誤，因為它是指在提供的叢集/無伺服器端點中不存在的 Amazon Redshift 資料庫。Amazon Redshift 無法驗證檢視，回應中的 `Status` 欄位`GetTable`會顯示錯誤 (`"State": "FAILED"`來自 Amazon Redshift。

GetTable 請求：

```
{
    "CatalogId": "123456789012",
    "DatabaseName": "async-view-test-db",
    "TableInput": {
        "Name": "view-athena-redshift-73",
        "Description": "This is an atomic operation",
        "StorageDescriptor": {
            "Columns": [
                { "Name": "col1", "Type": "int" },
                { "Name": "col2", "Type": "string" },
                { "Name": "col3", "Type": "double" }
            ]
        },
        "ViewDefinition": {
            "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
            "SubObjects": [ "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1" ],
            "Representations": [
                {
                    "Dialect": "ATHENA",
                    "DialectVersion": "3",
                    "ViewOriginalText": "SELECT * FROM \"gdc-view-playground-db\".
\"table_1\"",
                    "ValidationConnection": "athena-connection"
                },
                {
                    "Dialect": "REDSHIFT",
                    "DialectVersion": "1.0",
                    "ViewOriginalText": "SELECT * FROM \"gdc-view-playground-external-
schema\".\"table_1\";",
                    "ValidationConnection": "redshift-connection-malformed"
                }
            ]
        }
    }
}
```

GetTable 回應：

```
IncludeStatusDetails = FALSE
{
```

```
    "Table": {
        "Name": "view-athena-redshift-73",
        "DatabaseName": "async-view-test-db",
        "Description": "",
        "CreateTime": "2024-07-11T11:43:27-07:00",
        "UpdateTime": "2024-07-11T11:43:27-07:00",
        "Retention": 0,
        "ViewOriginalText": "",
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "IsMultiDialectView": false,
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:43:27-07:00",
            "UpdateTime": "2024-07-11T11:43:40-07:00",
            "Action": "CREATE",
            "State": "FAILED"
        }
    }
}

IncludeStatusDetails = TRUE
{
    "Table": {
        "Name": "view-athena-redshift-73",
        "DatabaseName": "async-view-test-db",
        "Description": "",
        "CreateTime": "2024-07-11T11:43:27-07:00",
        "UpdateTime": "2024-07-11T11:43:27-07:00",
        "Retention": 0,
        "ViewOriginalText": "",
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
```

```
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "IsMultiDialectView": false,
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:43:27-07:00",
            "UpdateTime": "2024-07-11T11:43:40-07:00",
            "Action": "CREATE",
            "State": "FAILED",
            "Error": {
                "ErrorCode": "QueryExecutionException",
                "ErrorMessage": "Error received during view SQL validation using a
connection: [Connection Name: redshift-connection-malformed | Query Execution Id:
69bfafd4-3d51-4cb0-9320-7ce5404b1809 | Timestamp: Thu Jul 11 18:43:38 UTC 2024]:
Redshift returned error for the statement: FATAL: database \"devooo\" does not exist"
            },
            "Details": {
                "RequestedChange": {
                    "Name": "view-athena-redshift-73",
                    "DatabaseName": "async-view-test-db",
                    "Description": "This is an atomic operation",
                    "Retention": 0,
                    "StorageDescriptor": {
                        "Columns": [
                            {
                                "Name": "col1",
                                "Type": "int"
                            },
                            {
                                "Name": "col2",
                                "Type": "string"
                            },
                            {
                                "Name": "col3",
                                "Type": "double"
                            }
                        ],
                        "Compressed": false,
                        "NumberOfBuckets": 0,
                        "SortColumns": [],
                        "StoredAsSubDirectories": false
                    },
                    "TableType": "VIRTUAL_VIEW",
```

```
                    "IsRegisteredWithLakeFormation": false,
                    "CatalogId": "123456789012",
                    "IsRowFilteringEnabled": false,
                    "VersionId": "-1",
                    "DatabaseId": "<databaseID>",
                    "ViewDefinition": {
                        "IsProtected": true,
                        "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
                        "SubObjects": [
                            "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1"
                        ],
                        "Representations": [
                            {
                                "Dialect": "ATHENA",
                                "DialectVersion": "3",
                                "ViewOriginalText": "SELECT * FROM \"gdc-view-
playground-db\".\"table_1\"",
                                "IsStale": false
                            },
                            {
                                "Dialect": "REDSHIFT",
                                "DialectVersion": "1.0",
                                "ViewOriginalText": "SELECT * FROM \"gdc-view-
playground-external-schema\".\"table_1\";",
                                "IsStale": false
                            }
                        ]
                    },
                    "IsMultiDialectView": true
                },
                "ViewValidations": [
                    {
                        "Dialect": "ATHENA",
                        "DialectVersion": "3",
                        "ViewValidationText": "SELECT * FROM \"gdc-view-playground-db
\".\"table_1\"",
                        "UpdateTime": "2024-07-11T11:43:40-07:00",
                        "State": "SUCCESS"
                    },
                    {
                        "Dialect": "REDSHIFT",
                        "DialectVersion": "1.0",
```

```
                        "ViewValidationText": "SELECT * FROM \"gdc-view-playground-
external-schema\".\"table_1\";",
                        "UpdateTime": "2024-07-11T11:43:38-07:00",
                        "State": "FAILED",
                        "Error": {
                            "ErrorCode": "QueryExecutionException",
                            "ErrorMessage": "Error received during view SQL validation
 using a connection: [Connection Name: redshift-connection-malformed | Query Execution
 Id: 69bfafd4-3d51-4cb0-9320-7ce5404b1809 | Time
stamp: Thu Jul 11 18:43:38 UTC 2024]: Redshift returned error for the statement: FATAL:
 database \"devooo\" does not exist"
                        }
                    }
                ]
            }
        }
    }
}
```

案例 3：Athena 查詢失敗

這裡的 SQL for Athena 無效，因為查詢漏接資料庫名稱。Athena 查詢驗證會擷取此項目，產生的錯誤會透過`GetTable`呼叫中的`Status`物件呈現。

`GetTable` 請求：

```
{
    "CatalogId": "123456789012",
    "DatabaseName": "async-view-test-db",
    "TableInput": {
        "Name": "view-athena-redshift-70",
        "Description": "This is an atomic operation",
        "StorageDescriptor": {
            "Columns": [
                { "Name": "col1", "Type": "int" },
                { "Name": "col2", "Type": "string" },
                { "Name": "col3", "Type": "double" }
            ]
        },
        "ViewDefinition": {
            "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
```

```
            "SubObjects": [ "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1" ],
            "Representations": [
                {
                    "Dialect": "ATHENA",
                    "DialectVersion": "3",
                    "ViewOriginalText": "SELECT * FROM \"gdc--view-playground-db\".
\"table_1\"",
                    "ValidationConnection": "athena-connection"
                },
                {
                    "Dialect": "REDSHIFT",
                    "DialectVersion": "1.0",
                    "ViewOriginalText": "SELECT * FROM \"gdc-view-playground-external-
schema\".\"table_1\";",
                    "ValidationConnection": "redshift-connection"
                }
            ]
        }
    }
}
```

GetTable 回應：

```
IncludeStatusDetails = FALSE
{
    "Table": {
        "Name": "view-athena-redshift-70",
        "DatabaseName": "async-view-test-db",
        "Description": "",
        "CreateTime": "2024-07-11T11:09:53-07:00",
        "UpdateTime": "2024-07-11T11:09:53-07:00",
        "Retention": 0,
        "ViewOriginalText": "",
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "IsMultiDialectView": false,
```

```
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:09:54-07:00",
            "UpdateTime": "2024-07-11T11:10:41-07:00",
            "Action": "CREATE",
            "State": "FAILED",
        }
    }
}

IncludeStatusDetails = TRUE
{
    "Table": {
        "Name": "view-athena-redshift-70",
        "DatabaseName": "async-view-test-db",
        "Description": "",
        "CreateTime": "2024-07-11T11:09:53-07:00",
        "UpdateTime": "2024-07-11T11:09:53-07:00",
        "Retention": 0,
        "ViewOriginalText": "",
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "IsMultiDialectView": false,
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:09:54-07:00",
            "UpdateTime": "2024-07-11T11:10:41-07:00",
            "Action": "CREATE",
            "State": "FAILED",
            "Error": {
                "ErrorCode": "QueryExecutionException",
                "ErrorMessage": "Error received during view SQL validation using
 a connection: [Connection Name: athena-connection | Query Execution Id: d9bb1e6d-
ce26-4b35-8276-8a199af966aa | Timestamp: Thu Jul 11 18:10:
41 UTC 2024]: Athena validation FAILED: {ErrorCategory: 2,ErrorType: 1301,Retryable:
 false,ErrorMessage: line 1:118: Schema 'gdc--view-playground-db' does not exist}"
```

```
            },
            "Details": {
                "RequestedChange": {
                    "Name": "view-athena-redshift-70",
                    "DatabaseName": "async-view-test-db",
                    "Description": "This is an atomic operation",
                    "Retention": 0,
                    "StorageDescriptor": {
                        "Columns": [
                            {
                                "Name": "col1",
                                "Type": "int"
                            },
                            {
                                "Name": "col2",
                                "Type": "string"
                            },
                            {
                                "Name": "col3",
                                "Type": "double"
                            }
                        ],
                        "Compressed": false,
                        "NumberOfBuckets": 0,
                        "SortColumns": [],
                        "StoredAsSubDirectories": false
                    },
                    "TableType": "VIRTUAL_VIEW",
                    "IsRegisteredWithLakeFormation": false,
                    "CatalogId": "123456789012",
                    "IsRowFilteringEnabled": false,
                    "VersionId": "-1",
                    "DatabaseId": "<databaseID>",
                    "ViewDefinition": {
                        "IsProtected": true,
                        "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
                        "SubObjects": [
                            "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1"
                        ],
                        "Representations": [
                            {
                                "Dialect": "ATHENA",
                                "DialectVersion": "3",
```

```
                                "ViewOriginalText": "SELECT * FROM \"gdc--view-
playground-db\".\"table_1\"",
                                "IsStale": false
                            },
                            {
                                "Dialect": "REDSHIFT",
                                "DialectVersion": "1.0",
                                "ViewOriginalText": "SELECT * FROM \"gdc-view-
playground-external-schema\".\"table_1\";",
                                "IsStale": false
                            }
                        ]
                    },
                    "IsMultiDialectView": true
                },
                "ViewValidations": [
                    {
                        "Dialect": "ATHENA",
                        "DialectVersion": "3",
                        "ViewValidationText": "SELECT * FROM \"gdc--view-playground-db
\".\"table_1\"",
                        "UpdateTime": "2024-07-11T11:10:41-07:00",
                        "State": "FAILED",
                        "Error": {
                            "ErrorCode": "QueryExecutionException",
                            "ErrorMessage": "Error received during view SQL validation
 using a connection: [Connection Name: athena-connection | Query Execution Id:
 d9bb1e6d-ce26-4b35-8276-8a199af966aa | Timestamp: Thu J
ul 11 18:10:41 UTC 2024]: Athena validation FAILED: {ErrorCategory: 2,ErrorType:
 1301,Retryable: false,ErrorMessage: line 1:118: Schema 'gdc--view-playground-db' does
 not exist}"
                        }
                    },
                    {
                        "Dialect": "REDSHIFT",
                        "DialectVersion": "1.0",
                        "ViewValidationText": "SELECT * FROM \"gdc-view-playground-
external-schema\".\"table_1\";",
                        "UpdateTime": "2024-07-11T11:10:41-07:00",
                        "State": "SUCCESS"
                    }
                ]
            }
        }
```

```
    }
}
```

案例 4：不相符的儲存描述項

Athena 方言提供的 SQL 會選取 `col1`，`col2`而 SQL for Redshift 只會選取 `col1`。這會導致儲存描述項不相符錯誤。

`GetTable` 請求：

```
{
    "CatalogId": "123456789012",
    "DatabaseName": "async-view-test-db",
    "TableInput": {
        "Name": "view-athena-redshift-71",
        "Description": "This is an atomic operation",
        "StorageDescriptor": {
            "Columns": [
                { "Name": "col1", "Type": "int" },
                { "Name": "col2", "Type": "string" },
                { "Name": "col3", "Type": "double" }
            ]
        },
        "ViewDefinition": {
            "Definer": "arn:aws:iam::123456789012:role/GDCViewDdefiner",
            "SubObjects": [ "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1" ],
            "Representations": [
                {
                    "Dialect": "ATHENA",
                    "DialectVersion": "3",
                    "ViewOriginalText": "SELECT col1, col2 FROM \"gdc-view-playground-
db\".\"table_1\"",
                    "ValidationConnection": "athena-connection"
                },
                {
                    "Dialect": "REDSHIFT",
                    "DialectVersion": "1.0",
                    "ViewOriginalText": "SELECT col1 FROM \"gdc-view-playground-
external-schema\".\"table_1\";",
                    "ValidationConnection": "redshift-connection"
                }
            ]
```

```
        }
    }
}
```

GetTable 回應：

```
IncludeStatusDetails = FALSE

{
    "Table": {
        "Name": "view-athena-redshift-71",
        "DatabaseName": "async-view-test-db",
        "Description": "",
        "CreateTime": "2024-07-11T11:22:02-07:00",
        "UpdateTime": "2024-07-11T11:22:02-07:00",
        "Retention": 0,
        "ViewOriginalText": "",
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "IsMultiDialectView": false,
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:22:02-07:00",
            "UpdateTime": "2024-07-11T11:23:19-07:00",
            "Action": "CREATE",
            "State": "FAILED"
        }
    }
}

IncludeStatusDetails = TRUE

{
    "Table": {
        "Name": "view-athena-redshift-71",
        "DatabaseName": "async-view-test-db",
```

```
        "Description": "",
        "CreateTime": "2024-07-11T11:22:02-07:00",
        "UpdateTime": "2024-07-11T11:22:02-07:00",
        "Retention": 0,
        "ViewOriginalText": "",
        "ViewExpandedText": "",
        "TableType": "",
        "CreatedBy": "arn:aws:iam::123456789012:user/zcaisse",
        "IsRegisteredWithLakeFormation": false,
        "CatalogId": "123456789012",
        "IsRowFilteringEnabled": false,
        "VersionId": "-1",
        "DatabaseId": "<databaseID>",
        "IsMultiDialectView": false,
        "Status": {
            "RequestedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "UpdatedBy": "arn:aws:iam::123456789012:user/zcaisse",
            "RequestTime": "2024-07-11T11:22:02-07:00",
            "UpdateTime": "2024-07-11T11:23:19-07:00",
            "Action": "CREATE",
            "State": "FAILED",
            "Error": {
                "ErrorCode": "InvalidInputException",
                "ErrorMessage": "Engine and existing storage descriptor mismatch"
            },
            "Details": {
                "RequestedChange": {
                    "Name": "view-athena-redshift-71",
                    "DatabaseName": "async-view-test-db",
                    "Description": "This is an atomic operation",
                    "Retention": 0,
                    "StorageDescriptor": {
                        "Columns": [
                            {
                                "Name": "col1",
                                "Type": "int"
                            },
                            {
                                "Name": "col2",
                                "Type": "string"
                            },
                            {
                                "Name": "col3",
                                "Type": "double"
```

```
                            }
                        ],
                        "Compressed": false,
                        "NumberOfBuckets": 0,
                        "SortColumns": [],
                        "StoredAsSubDirectories": false
                    },
                    "TableType": "VIRTUAL_VIEW",
                    "IsRegisteredWithLakeFormation": false,
                    "CatalogId": "123456789012",
                    "IsRowFilteringEnabled": false,
                    "VersionId": "-1",
                    "DatabaseId": "<databaseID>",
                    "ViewDefinition": {
                        "IsProtected": true,
                        "Definer": "arn:aws:iam::123456789012:role/GDCViewDefiner",
                        "SubObjects": [
                            "arn:aws:glue:us-east-1:123456789012:table/gdc-view-
playground-db/table_1"
                        ],
                        "Representations": [
                            {
                                "Dialect": "ATHENA",
                                "DialectVersion": "3",
                                "ViewOriginalText": "SELECT col1, col2 FROM \"gdc-view-
playground-db\".\"table_1\"",
                                "IsStale": false
                            },
                            {
                                "Dialect": "REDSHIFT",
                                "DialectVersion": "1.0",
                                "ViewOriginalText": "SELECT col1 FROM \"gdc-view-
playground-external-schema\".\"table_1\";",
                                "IsStale": false
                            }
                        ]
                    },
                    "IsMultiDialectView": true
                },
                "ViewValidations": [
                    {
                        "Dialect": "ATHENA",
                        "DialectVersion": "3",
```

```
                        "ViewValidationText": "SELECT col1, col2 FROM \"gdc-view-
playground-db\".\"table_1\"",
                        "UpdateTime": "2024-07-11T11:23:19-07:00",
                        "State": "FAILED",
                        "Error": {
                            "ErrorCode": "InvalidInputException",
                            "ErrorMessage": "Engine and existing storage descriptor
 mismatch"
                        }
                    },
                    {
                        "Dialect": "REDSHIFT",
                        "DialectVersion": "1.0",
                        "ViewValidationText": "SELECT col1 FROM \"gdc-view-playground-
external-schema\".\"table_1\";",
                        "UpdateTime": "2024-07-11T11:22:49-07:00",
                        "State": "FAILED",
                        "Error": {
                            "ErrorCode": "InvalidInputException",
                            "ErrorMessage": "Engine and existing storage descriptor
 mismatch"
                        }
                    }
                ]
            }
        }
    }
}
```

授予 Data Catalog 檢視的許可

在 中建立檢視後 AWS Glue Data Catalog，您可以將檢視的資料湖許可授予 AWS 帳戶、組織和組織單位的主體。您可以使用 LF-Tags 或具名資源方法授予許可。如需標記資源的詳細資訊，請參閱 Lake Formation 標籤型存取控制。如需直接授予檢視許可的詳細資訊，請參閱 使用具名資源方法授予檢視的許可 。

在 Lake Formation 中使用工作流程匯入資料

使用 AWS Lake Formation，您可以使用工作流程匯入資料。工作流程會定義將資料匯入資料湖的資料來源和排程。這是AWS Glue爬蟲程式、任務和觸發程序的容器，用於協調程序以載入和更新資料湖。

主題

- Lake Formation 中的藍圖和工作流程
- 建立工作流程
- 執行工作流程

Lake Formation 中的藍圖和工作流程

工作流程會封裝複雜的多工作業擷取、轉換和載入 (ETL) 活動。工作流程會產生 AWS Glue 爬蟲程式、任務和觸發程序，以協調資料的載入和更新。Lake Formation 會以單一實體的形式執行和追蹤工作流程。您可以設定工作流程以隨需或排程執行。

> **Note**
>
> Spark parquet 寫入器不支援資料欄名稱中的特殊字元。這是寫入器本身的技術限制，而不是組態問題。

您在 Lake Formation 中建立的工作流程會在AWS Glue主控台中顯示為導向無環圖 (DAG)。每個 DAG 節點都是任務、爬蟲程式或觸發程序。若要監控進度和疑難排解，您可以追蹤工作流程中每個節點的狀態。

當 Lake Formation 工作流程完成時，執行工作流程的使用者會獲得工作流程所建立之 Data Catalog 資料表的 Lake Formation SELECT許可。

您也可以在 中建立工作流程AWS Glue。不過，由於 Lake Formation 可讓您從藍圖建立工作流程，因此在 Lake Formation 中建立工作流程更為簡單且自動化。Lake Formation 提供下列類型的藍圖：

- 資料庫快照 – 從 JDBC 來源將資料從所有資料表載入或重新載入至資料湖。您可以根據排除模式從來源排除一些資料。
- 增量資料庫 – 根據先前設定的書籤，僅從 JDBC 來源將新資料載入資料湖。您可以在 JDBC 來源資料庫中指定要包含的個別資料表。對於每個資料表，您可以選擇書籤資料欄和書籤排序順序，以追蹤先前載入的資料。第一次針對一組資料表執行增量資料庫藍圖時，工作流程會從資料表載入所有資料，並為下一個增量資料庫藍圖執行設定書籤。因此，您可以使用增量資料庫藍圖而非資料庫快照藍圖來載入所有資料，前提是您將資料來源中的每個資料表指定為參數。
- 日誌檔案 – 從日誌檔案來源大量載入資料 AWS CloudTrail，包括 Elastic Load Balancing 日誌和 Application Load Balancer 日誌。

使用下表來協助決定是否要使用資料庫快照或增量資料庫藍圖。

使用資料庫快照時...	在下列情況下使用增量資料庫...
• 結構描述演變具有彈性。（已重新命名資料欄，刪除先前的資料欄，並在其位置新增資料欄。） • 來源和目的地之間需要完整的一致性。	• 結構描述演變是增量的。（只有連續新增的資料欄。） • 只會新增新資料列；先前的資料列不會更新。

> **Note**
>
> 使用者無法編輯 Lake Formation 建立的藍色列印和工作流程。

建立工作流程

開始之前，請確定您已將必要的資料許可和資料位置許可授予角色 `LakeFormationWorkflowRole`。這是為了讓工作流程可以在 Data Catalog 中建立中繼資料表，並將資料寫入 Amazon S3 中的目標位置。如需詳細資訊，請參閱（選用）建立工作流程的 IAM 角色及Lake Formation 許可概觀 。

> **Note**
>
> Lake Formation 使用 `GetTemplateInstance`、 `GetTemplateInstances`和 `InstantiateTemplate`操作從藍圖建立工作流程。這些操作不可公開使用，且僅用於內部代表您建立資源。您會收到用於建立工作流程的 CloudTrail 事件。

從藍圖建立工作流程

1. 在 https：//https://console.aws.amazon.com/lakeformation/ 開啟 AWS Lake Formation 主控台。以資料湖管理員或具有資料工程師許可的使用者身分登入。如需詳細資訊，請參閱Lake Formation 角色和 IAM 許可參考。
2. 在導覽窗格中，選擇藍圖，然後選擇使用藍圖。
3. 在使用藍圖頁面上，選擇圖磚以選取藍圖類型。
4. 在匯入來源下，指定資料來源。

   如果您是從 JDBC 來源匯入，請指定下列項目：

- 資料庫連線 – 從清單中選擇連線。使用 AWS Glue主控台建立其他連線。連線中的 JDBC 使用者名稱和密碼會決定工作流程可存取的資料庫物件。
- 來源資料路徑 – 輸入 *<database>*/*<schema>*/*<table>* 或 *<database>*/*<table>*，視資料庫產品而定。Oracle 資料庫和 MySQL 不支援路徑中的結構描述。您可以用百分比 (%) 字元取代 *<schema>* 或 *<table>*。例如，對於系統識別符 (SID) 為 的 Oracle 資料庫`orcl`，輸入 `orcl/%` 以匯入連線中名為 之使用者可存取的所有資料表。

  > **Important**
  >
  > 此欄位區分大小寫。如果任何元件的案例不相符，工作流程將會失敗。

  如果您指定 MySQL 資料庫， AWS Glue ETL 預設會使用 Mysql5 JDBC 驅動程式，因此原生不支援 MySQL8。您可以編輯 ETL 任務指令碼來使用`customJdbcDriverS3Path`參數，如《 AWS Glue 開發人員指南》中的 JDBC connectionType 值中所述，以使用支援 MySQL8 的不同 JDBC 驅動程式。

如果您是從日誌檔案匯入 ，請確定您為工作流程指定的角色 (「工作流程角色」) 具有存取資料來源所需的 IAM 許可。例如，若要匯入 AWS CloudTrail 日誌，使用者必須具有`cloudtrail:DescribeTrails`和 `cloudtrail:LookupEvents`許可，才能在建立工作流程時查看 CloudTrail 日誌清單，而且工作流程角色必須具有 Amazon S3 中 CloudTrail 位置的許可。

5. 執行以下任意一項：

   - 對於資料庫快照藍圖類型，選擇性地指定一或多個排除模式來識別要匯入的資料子集。這些排除模式是 Unix 樣式`glob`模式。它們會儲存為工作流程所建立資料表的屬性。

     如需可用排除模式的詳細資訊，請參閱《 AWS Glue 開發人員指南》中的包含和排除模式。
   - 針對增量資料庫藍圖類型，指定下列欄位。為每個要匯入的資料表新增一列。

     資料表名稱

     要匯入的資料表。必須是全部小寫。

     書籤索引鍵

     定義書籤索引鍵的資料欄名稱逗號分隔清單。如果為空白，則會使用主索引鍵來判斷新資料。每個資料欄的案例必須符合資料來源中定義的案例。

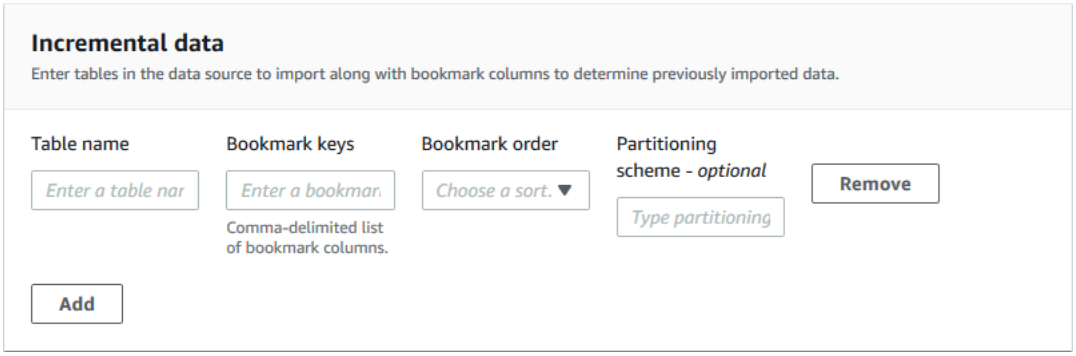
> **Note**
>
> 主索引鍵只有在循序增加或減少（沒有間隙）時，才符合預設書籤索引鍵的資格。如果您想要使用主索引鍵做為書籤索引鍵且具有間隙，則必須將主索引鍵欄命名為書籤索引鍵。

書籤順序

當您選擇遞增時，值大於書籤值的資料列會識別為新資料列。當您選擇遞減時，值小於書籤值的資料列會識別為新資料列。

分割結構

（選用）分割索引鍵資料欄的清單，以斜線 (/) 分隔。範例： `year/month/day`。



如需詳細資訊，請參閱《 AWS Glue 開發人員指南》中的使用任務書籤追蹤已處理的資料。

6. 在匯入目標下，指定目標資料庫、目標 Amazon S3 位置和資料格式。

   確定工作流程角色對資料庫和 Amazon S3 目標位置具有必要的 Lake Formation 許可。

   > **Note**
   >
   > 目前，藍圖不支援加密目標的資料。

7. 選擇匯入頻率。

   您可以使用自訂選項指定`cron`表達式。

8. 在匯入選項下：

   a. 輸入工作流程名稱。

b. 針對角色，選擇LakeFormationWorkflowRole您在 中建立的角色（選用）建立工作流程的 IAM 角色。

c. 選擇性地指定資料表字首。字首會加上工作流程建立的 Data Catalog 資料表名稱。

9. 選擇建立，然後等待主控台報告工作流程已成功建立。

> **Tip**
>
> 您是否收到下列錯誤訊息？
> `User: arn:aws:iam::<account-id>:user/<username> is not authorized to perform: iam:PassRole on resource:arn:aws:iam::<account-id>:role/<rolename>...`
> 若是如此，請檢查您是否在所有政策中將 *<account-id>* 取代為有效的 AWS 帳號。

> **另請參閱：**
>
> - Lake Formation 中的藍圖和工作流程

執行工作流程

您可以使用 Lake Formation 主控台、AWS Glue主控台或AWS Glue命令列界面 (AWS CLI) 或 API 來執行工作流程。

執行工作流程 (Lake Formation 主控台）

1. 在 https：//https://console.aws.amazon.com/lakeformation/ 開啟 AWS Lake Formation 主控台。以資料湖管理員或具有資料工程師許可的使用者身分登入。如需詳細資訊，請參閱Lake Formation 角色和 IAM 許可參考。
2. 在導覽窗格中，選擇 Blueprints (藍圖)。
3. 在藍圖頁面上，選取工作流程。然後在動作功能表上，選擇開始。
4. 當工作流程執行時，請在上次執行狀態欄中檢視其進度。偶爾選擇重新整理按鈕。

   狀態會從 RUNNING 到 Discovering、匯入，再到 COMPLETED。

   當工作流程完成時：

- Data Catalog 有新的中繼資料表。
- 您的資料會擷取至資料湖。

如果工作流程失敗，請執行下列動作：

a. 選取工作流程。選擇動作，然後選擇檢視圖形。

   工作流程會在 AWS Glue主控台中開啟。

b. 確認已選取工作流程，然後選擇 History (歷史記錄) 標籤。

c. 在歷史記錄下，選取最近的執行，然後選擇檢視執行詳細資訊。

d. 在動態（執行時間）圖形中選取失敗的任務或爬蟲程式，然後檢閱錯誤訊息。失敗的節點為紅色或黃色。

另請參閱：

- Lake Formation 中的藍圖和工作流程

將您的資料帶入 AWS Glue Data Catalog

您可以在 AWS Glue Data Catalog （資料目錄） 中建立聯合目錄，並在 Amazon S3 資料湖和 Amazon Redshift 資料倉儲中統一資料。您也可以整合操作資料庫中的資料，例如 PostgreSQL Amazon DynamoDB、Google BigQuery、MySQL 等第三方資料來源。Data Catalog 提供集中式中繼資料儲存庫，可讓您更輕鬆地管理和探索不同系統中的資料。

Data Catalog 透過聯合連接器與超過 30 個外部資料來源整合。透過此整合，您可以查詢來自這些外部來源的資料，而不必建置資料管道來 AWS 先將資料擷取至 。

為外部資料編製目錄之後，您可以使用 AWS Lake Formation 在 Data Catalog 中集中管理資料存取許可。資料湖管理員可以將精細存取許可授予相同帳戶或跨帳戶的其他 IAM 主體 （使用者或角色）。IAM 主體接著可以使用各種 AWS 服務查詢資料，例如 Athena、Amazon EMR 或 Redshift Spectrum。

Data Catalog 提供下列方法來管理外部資料集和外部中繼存放區的資料和許可：

- 將 Amazon Redshift 資料倉儲中的資料帶入 AWS Glue Data Catalog：向 Data Catalog 註冊現有的 Amazon Redshift 命名空間或叢集，並在 Data Catalog 中建立多層聯合目錄。

  您可以使用任何與 Apache Iceberg REST 目錄 OpenAPI 規格相容的查詢引擎來存取資料，例如 Amazon EMR Serverless 和 Amazon Athena。
- 從外部資料來源聯合到 Data Catalog – 使用連線將 Data Catalog 連接到外部資料來源 AWS Glue ，並使用 Lake Formation 建立聯合目錄以集中管理資料集的存取許可。不需要將中繼資料遷移至 Data Catalog。
- 將 Amazon S3 Table 儲存貯體與 Data Catalog （預覽） 整合 – 您可以將 Amazon S3 Tables 發佈和分類為 Data Catalog 物件，並從 Lake Formation 主控台或使用 AWS Glue API 操作將目錄註冊為 Lake Formation 資料位置。
- 建立目錄以在 Data Catalog 中管理 Amazon Redshift 資料表 – 您目前可能沒有可用的 Amazon Redshift 生產者叢集或 Amazon Redshift 資料共用，但想要使用 Data Catalog 建立和管理 Amazon Redshift 資料表。您可以使用 `glue:CreateCatalog` API 操作或 AWS Lake Formation 主控台建立 AWS Glue 受管目錄，將目錄類型設定為 `Managed`和 `Catalog source` Redshift，以開始使用。
- 使用 Data Catalog 發佈 Amazon Redshift 資料共用 – 將 Amazon Redshift 資料共用發佈至 Data Catalog，並使用 Lake Formation 集中管理資料共用的資料存取並限制使用者存取。

  您可以使用 Amazon Redshift Spectrum 查詢資料。

- 將 Data Catalog 連接到外部 Hive 中繼存放區 – 將 Data Catalog 連接到外部中繼存放區，以使用 Lake Formation 管理 Amazon S3 中資料集的存取許可。不需要將中繼資料遷移至 Data Catalog。
- 整合 Lake Formation 與 AWS 資料交換 – Lake Formation 支援透過 授權存取您的資料 AWS Data Exchange。如果您想要授權 Lake Formation 資料，請參閱AWS Data Exchange 《 使用者指南》中的什麼是 AWS Data Exchange 。

主題

- 將 Amazon Redshift 資料帶入 AWS Glue Data Catalog
- 在 中聯合到外部資料來源 AWS Glue Data Catalog
- 在 中建立 Amazon S3 Tables 目錄 AWS Glue Data Catalog
- 在 中建立 Amazon Redshift 受管目錄 AWS Glue Data Catalog
- 管理 Amazon Redshift 資料共用中資料的許可
- 管理使用外部中繼存放區的資料集許可

將 Amazon Redshift 資料帶入 AWS Glue Data Catalog

您可以在 AWS Glue Data Catalog （資料目錄） 中管理 Amazon Redshift 資料倉儲中的分析資料，並統一 Amazon S3 資料湖和 Amazon Redshift 資料倉儲。Amazon Redshift 是 AWS 雲端中全受管的 PB 級資料倉儲服務。Amazon Redshift 資料倉儲是稱為節點的運算資源的集合，組織成稱為叢集的群組。每個叢集皆執行 Amazon Redshift 引擎並包含一或多個資料庫。

在 Amazon Redshift 中，您可以建立 Amazon Redshift 佈建叢集和無伺服器命名空間，並將其註冊至 Data Catalog。透過這樣做，您可以統一 Amazon Redshift 受管儲存 (RMS) 和 Amazon S3 儲存貯體中的資料，並從 Apache Iceberg 相容分析引擎存取資料。

透過註冊命名空間和叢集，您可以提供對資料的存取，而不需要複製或移動資料。如需在 Amazon Redshift 中註冊叢集和命名空間的詳細資訊，請參閱向 註冊 Amazon Redshift 叢集和命名空間 AWS Glue Data Catalog。

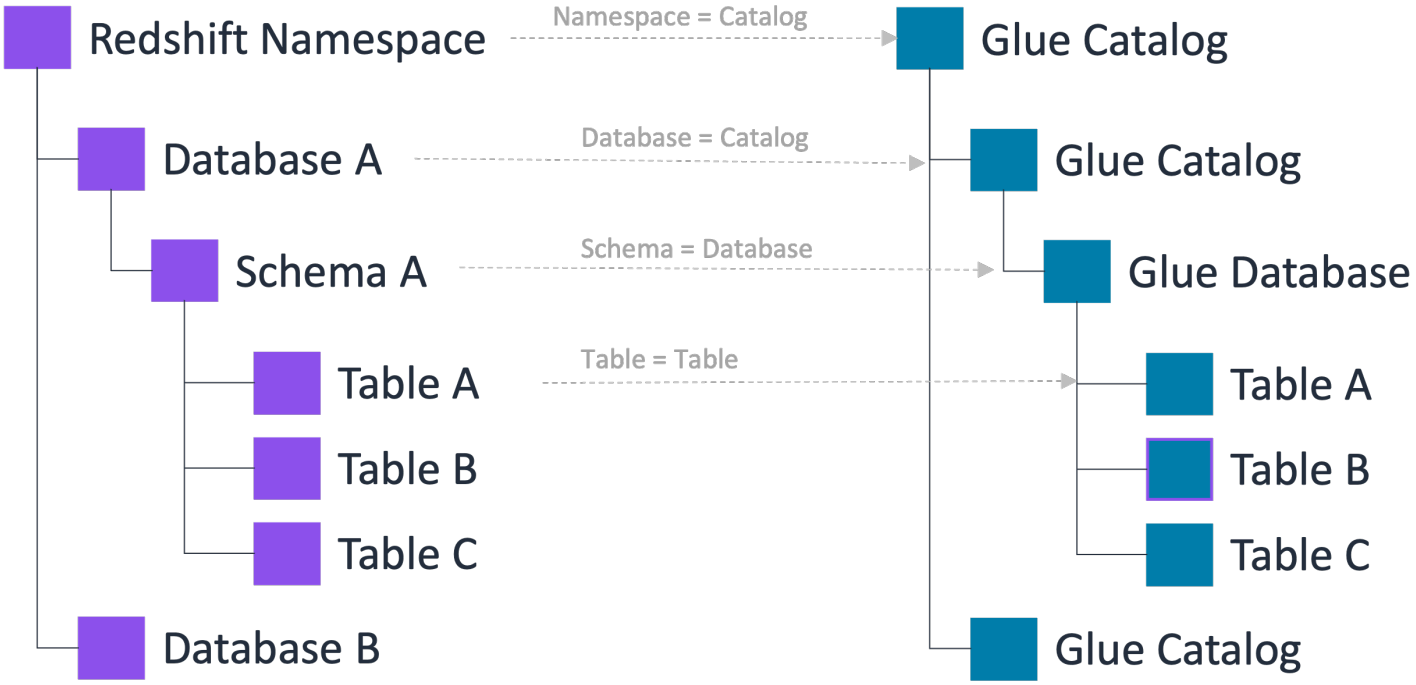
在 Amazon Redshift 中，您可以透過資料共用或向 Data Catalog 註冊命名空間和叢集來執行資料共用。使用在個別資料庫物件層級操作的資料共用，您必須為每個資料表或檢視啟用共用。相反地，命名空間會在叢集或命名空間層級發佈函數。當您向 Data Catalog 註冊叢集或命名空間時，它中的所有資料庫和資料表都會自動共用，而不必為個別物件設定共用。

在資料目錄中，您可以為每個命名空間或叢集建立聯合目錄。當目錄指向 Data Catalog 外部的實體時，稱為聯合目錄。Amazon Redshift 命名空間中的資料表和檢視會列為 Data Catalog 中的個別資料

表。您可以與同一帳戶內或 Lake Formation 的其他帳戶中的選定 IAM 主體和 SAML 使用者共用聯合目錄中的資料庫和資料表。您也可以包含資料列和資料欄篩選條件表達式，以限制對特定資料的存取。如需詳細資訊，請參閱Lake Formation 中的資料篩選和儲存格層級安全性。

Data Catalog 支援包含目錄、資料庫和資料表 （和檢視） 的三層中繼資料階層。當您向 Data Catalog 註冊命名空間時，Amazon Redshift 資料階層會對應至 Data Catalog 的 3 層階層，如下所示：

- Amazon Redshift 命名空間會成為 Data Catalog 中的多層級目錄。
- 相關聯的 Amazon Redshift 資料庫會在 Data Catalog 中註冊為目錄。
- Amazon Redshift 結構描述會成為 Data Catalog 中的資料庫。
- Amazon Redshift 資料表會成為 Data Catalog 中的資料表。



透過此三層中繼資料階層，您可以使用 Data Catalog 中的 3 部分表示法 - "catalog1/catalog2.database.table" 來存取 Amazon Redshift 資料表。此外，資料團隊可以維護 Amazon Redshift 用來組織 Data Catalog 帳戶中資料表的相同組織。

在 Lake Formation 中，您可以使用 Data Catalog 資源的精細存取控制，安全地管理來自 Amazon Redshift 的資料。透過此整合，您可以使用常見的存取控制機制，管理、保護和查詢來自單一目錄的分析資料。

如需限制的詳細資訊，請參閱將 Amazon Redshift 資料倉儲資料帶入 的限制 AWS Glue Data Catalog。

主題

- 主要優點
- 角色和責任
- 在 中管理 Amazon Redshift 命名空間的先決條件 AWS Glue Data Catalog
- 建立 Amazon Redshift 聯合目錄
- 檢視目錄物件
- 更新聯合目錄
- 存取共用聯合目錄
- 刪除聯合目錄
- 查詢聯合目錄
- 其他資源

主要優點

使用 註冊 Amazon Redshift 叢集和命名空間， AWS Glue Data Catalog 並在 Amazon S3 資料湖和 Amazon Redshift 資料倉儲中統一資料，可提供下列優點：

- 統一查詢體驗 – 使用任何與 Apache Iceberg 相容的查詢引擎來查詢 Amazon Redshift 受管資料和 Amazon S3 儲存貯體中的資料，例如 Amazon EMR Serverless 和 Amazon Athena，而無需移動或複製資料。
- 跨服務一致的資料存取 – 從不同的 AWS 分析服務存取相同的聯合資料來源時，您不需要更新資料管道中的資料庫和資料表名稱，因為資料來源已在 Data Catalog 中註冊。
- 精細存取控制 – 您可以使用精細存取控制許可，套用 Lake Formation 許可來管理對聯合資料來源的存取。

角色和責任

Role	責任
Amazon Redshift 生產者叢集管理員	向 Data Catalog 註冊叢集或命名空間。

Lake Formation 資料湖管理員	接受叢集或命名空間邀請、建立聯合目錄，以及將聯合目錄的存取權授予其他主體。
Lake Formation 唯讀管理員	探索聯合目錄，查詢聯合目錄中的 Amazon Redshift 資料表。
資料傳輸角色	Amazon Redshift 會代表您在 Amazon S3 儲存貯體之間傳輸資料。

以下是提供使用者存取 Amazon Redshift 命名空間的高階步驟：

1. 在 Amazon Redshift 中，生產者叢集管理員向 Data Catalog 註冊叢集或命名空間。
2. 資料湖管理員接受來自 Amazon Redshift 生產者叢集管理員的命名空間邀請，並在 Data Catalog 中建立聯合目錄。

   完成此步驟後，您可以在 Data Catalog 中管理 Amazon Redshift 命名空間目錄。
3. 將許可授予目錄、資料庫和資料表上的使用者。您可以與相同帳戶或其他帳戶中的使用者共用整個命名空間目錄或資料表子集。

在 中管理 Amazon Redshift 命名空間的先決條件 AWS Glue Data Catalog

1. 建立資料湖管理員 - 建立有權接受命名空間邀請的 IAM 角色，並建立 AWS Glue Data Catalog 物件（目錄、資料庫、資料表/檢視），並將 Lake Formation 許可授予其他使用者。

   如需建立資料湖管理員的step-by-step說明，請參閱建立資料湖管理員。
2. 更新資料湖管理員許可。

   除了資料湖管理員許可之外，資料湖管理員還需要下列許可，才能在 Lake Formation 中接受 Amazon Redshift 命名空間邀請、建立或更新 Data Catalog 資源，以及啟用資料湖存取：

```
{
    "Version": "2012-10-17",
    "Id": "glue-enable-datalake-access",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "redshift:AssociateDataShareConsumer",
```

```
                "redshift:DescribeDataSharesForConsumer",
                "redshift:DescribeDataShares",
                "redshift-serverless:CreateNamespace",
                "redshift-serverless:CreateWorkgroup",
                "redshift-serverless:DeleteNamespace",
                "redshift-serverless:DeleteWorkgroup",
                "ec2:DescribeAccountAttributes",
                "ec2:DescribeSubnets",
                "ec2:DescribeSecurityGroups",
                "ec2:DescribeAvailabilityZones",
                "s3:createBucket",
                "s3:deleteBucket",
                "s3:putBucketPolicy",
                "s3:putEncryptionConfiguration",
                "s3:putLifecycleConfiguration",
                "s3:putBucketVersioning",
                "iam:CreateRole"
            ],
            "Resource": "*"
        }
    ]
}

{
    "Action": [
        "iam:PassRole"
    ],
    "Effect": "Allow",
    "Resource": "arn:aws:iam::*:role/data transfer role name",
    "Condition": {
        "StringLike": {
            "iam:PassedToService": [
                "glue.amazonaws.com"
            ]
        }
    }
}
```

3. 如果用於建立聯合目錄的 IAM 角色不是資料湖管理員，您需要授予該角色`Create catalog`許可。

   建立目錄建立者

   a. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

b. 在管理下選擇管理角色和任務。

c. 選擇 Grant (授予)。

d. 在授予許可畫面上，選擇 IAM 使用者或角色。

e. 選取建立目錄許可。

f. 或者，您也可以授予可授予的建立目錄許可。可授予許可允許目錄建立者將`Create catalog`許可授予其他主體。

g. 選擇 Grant (授予)。

AWS CLI 授予建立聯合目錄許可的範例。

```
aws lakeformation grant-permissions \
--cli-input-json \
'{
    "Principal": {
     "DataLakePrincipalIdentifier":"arn:aws:iam::123456789012:role/Admin"
    },
    "Resource": {
        "Catalog": {
        }
    },
    "Permissions": [
        "CREATE_CATALOG",
        "DESCRIBE"
    ]
}'
```

4. 建立唯讀管理員角色，以在 Amazon Redshift 查詢編輯器 v2 的資料目錄中探索 Amazon Redshift 聯合目錄。

若要從 Amazon Redshift 查詢編輯器 v2 查詢聯合目錄中的 Amazon Redshift 資料表，請確定唯讀管理員角色政策包含 Amazon Redshift 服務連結角色的 ARN`AWSServiceRoleForRedshift`。

```
 aws lakeformation put-data-lake-settings
        --region us-east-1 \
        --data-lake-settings \
 '{
   "DataLakeAdmins": [{"DataLakePrincipalIdentifier":"arn:aws:iam::123456789012:role/
Admin"}],
```

```
   "ReadOnlyAdmins":[{"DataLakePrincipalIdentifier":"arn:aws:iam::123456789012:role/
aws-service-role/redshift.amazonaws.com/AWSServiceRoleForRedshift"}],
   "CreateDatabaseDefaultPermissions":[],
   "CreateTableDefaultPermissions":[],
   "Parameters":{"CROSS_ACCOUNT_VERSION":"4","SET_CONTEXT":"TRUE"}
  }'
```

5. 建立 Amazon Redshift 可代表您擔任的資料傳輸角色，以往返 Amazon S3 儲存貯體傳輸資料。

   當您啟用 Athena、Amazon EMR on Amazon EC2 等 Apache Iceberg 相容查詢引擎的資料湖存取以存取 Data Catalog 中的 Amazon Redshift 資源時，您需要建立具有必要許可的 IAM 角色，才能執行往返 Amazon S3 儲存貯體的資料傳輸。

```
{
    "Version": "2012-10-17",
    "Id": "glue-enable-datalake-access",
    "Statement": [{
      "Sid": "DataTransferRolePolicy",
        "Effect": "Allow",
        "Action": [ "glue:GetCatalog",
                    "glue:GetDatabase",
                    "kms:GenerateDataKey",
                    "kms:Decrypt"],
        "Resource": "*"
    }
    ]
}
```

6. 將下列信任政策新增至 AWS Glue 和 Amazon Redshift 服務的資料傳輸角色，以擔任角色在 Amazon S3 儲存貯體之間傳輸資料。

```
        {
    "Version": "2012-10-17",
    "Statement": [{
        "Effect": "Allow",
        "Principal": {
        "Service": [
        "redshift.amazonaws.com",
        "glue.amazonaws.com"
        ]
```

```
    },
        "Action": "sts:AssumeRole"
    }]
}
```

7. 如果您使用客戶受管金鑰來加密 Amazon Redshift 叢集/命名空間中的資料，請將下列金鑰政策新增至 AWS KMS 金鑰。將帳號取代為有效的 AWS 帳號，並指定資料傳輸角色名稱。根據預設，Amazon Redshift 叢集中的資料會使用 KMS 金鑰加密。Lake Formation 提供建立自訂 KMS 金鑰以進行加密的選項。如果您使用客戶受管金鑰，則必須將特定金鑰政策新增至金鑰。

   如需管理客戶受管金鑰許可的詳細資訊，請參閱客戶受管金鑰。

```
{
    "Version": "2012-10-17",
    "Id": "auto-redshift-3",
    "Statement": [
        {
            "Sid": "Allow access through RedShift for all principals in the account
 that are authorized to use RedShift",
            "Effect": "Allow",
            "Principal": {
                "AWS": "*"
           },
            "Action": [
                "kms:Encrypt",
                "kms:Decrypt",
                "kms:ReEncrypt*",
                "kms:GenerateDataKey*",
                "kms:CreateGrant",
                "kms:DescribeKey"
            ],
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "kms:CallerAccount": "123456789012",
                    "kms:ViaService": "redshift.us-east-1.amazonaws.com"
                }
            }
        },
        {
        "Sid": "Allow access through RedShift-Serverless for all principals in the
 account that are authorized to use RedShift-Serverless",
```

```
        "Effect": "Allow",
        "Principal": {
            "AWS": "*"
        },
        "Action": [
            "kms:Encrypt",
            "kms:Decrypt",
            "kms:ReEncrypt*",
            "kms:GenerateDataKey*",
            "kms:CreateGrant",
            "kms:DescribeKey"
        ],
        "Resource": "*",
        "Condition": {
            "StringEquals": {
                "kms:CallerAccount": "123456789012",
                "kms:ViaService": "redshift-serverless.us-east-1.amazonaws.com"
            }
        }
        },
        {
            "Sid": "Allow direct access to key metadata to the account",
            "Effect": "Allow",
            "Principal": {
                "AWS": "arn:aws:iam::123456789012:root"
            },
            "Action": [
                "kms:Describe*",
                "kms:Get*",
                "kms:List*",
                "kms:RevokeGrant"
            ],
            "Resource": "*"
        },
       {
            "Sid": "Allow GenerateDataKey + Decrypt to the DataTransferRole via s3",
            "Effect": "Allow",
            "Principal": {
                "AWS": "arn:aws:iam::123456789012 :role/data-transfer-role-name"
            },
            "Action": [
                "kms:GenerateDataKey",
                "kms:Decrypt"
            ],
```

```
            "Resource": "*"
        },
        "Condition": {
            "StringEquals": {
                "kms:ViaService": "s3.us-east-1.amazonaws.com"
            }
        }
    ]
}
```

建立 Amazon Redshift 聯合目錄

本主題說明接受叢集或命名空間邀請、建立聯合多層級目錄，以及將許可授予其他主體所需的步驟。您可以使用 Lake Formation 主控台、 AWS Command Line Interface (AWS CLI) 或 APIs/SDKs來完成這些任務。本主題中的範例顯示同一帳戶中的生產者叢集/命名空間、資料目錄和資料取用者。

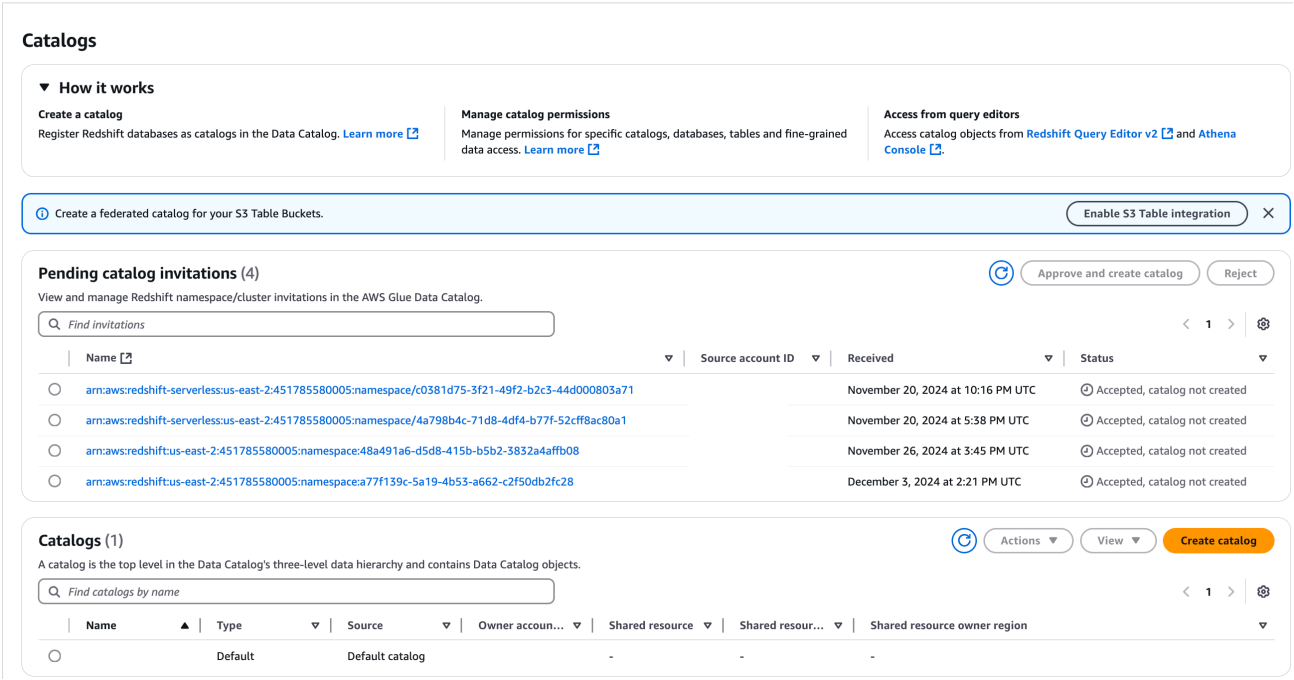
若要進一步了解 Lake Formation 跨帳戶功能，請參閱 Lake Formation 中的跨帳戶資料共用。

在 Data Catalog 中管理 Amazon Redshift 命名空間

1. 檢閱命名空間邀請並接受它。

   Console

   1. 以資料湖管理員身分登入 Lake Formation 主控台，網址為 https：//https://console.aws.amazon.com/lakeformation/。導覽至資料目錄下的目錄頁面。
   2. 檢閱您獲授權存取的命名空間邀請。狀態欄指出命名空間目前的參與狀態。未接受狀態表示您已新增至命名空間，但您尚未接受邀請或已拒絕邀請。



3. 若要回應命名空間或叢集邀請，請選取邀請名稱，然後選擇檢閱邀請。在接受或拒絕邀請中，檢閱邀請詳細資訊。選擇接受以接受邀請或拒絕以拒絕邀請。如果您拒絕邀請，則無法存取 命名空間。

AWS CLI

下列範例示範如何檢視、接受和註冊邀請。以有效的 AWS 帳戶 ID 取代 AWS 帳戶 ID。將 取代`data-share-arn`為參考命名空間的實際 Amazon Resource Name (ARN)。

1. 檢視擱置中的邀請。

```
aws redshift describe-data-shares \
 --data-share-arn 'arn:aws:redshift:us-
east-1:123456789012:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
ds_internal_namespace' \
```

2. 接受邀請。

```
 aws redshift associate-data-share-consumer \
 --data-share-arn 'arn:aws:redshift:us-
east-1:123456789012:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
ds_internal_namespace' \
 --consumer-arn 'arn:aws:glue:us-east-1:123456789012:catalog'
```

3. 在 Lake Formation 帳戶中註冊叢集或命名空間。使用 RegisterResource API 操作在 Lake Formation 中註冊資料共用。 `DataShareArn`是 的輸入參數`ResourceArn`。

> **Note**
>
> 這是強制性步驟。

```
aws lakeformation register-resource \
 --resource-arn 'arn:aws:redshift:us-
east-1:123456789012:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
ds_internal_namespace'
```

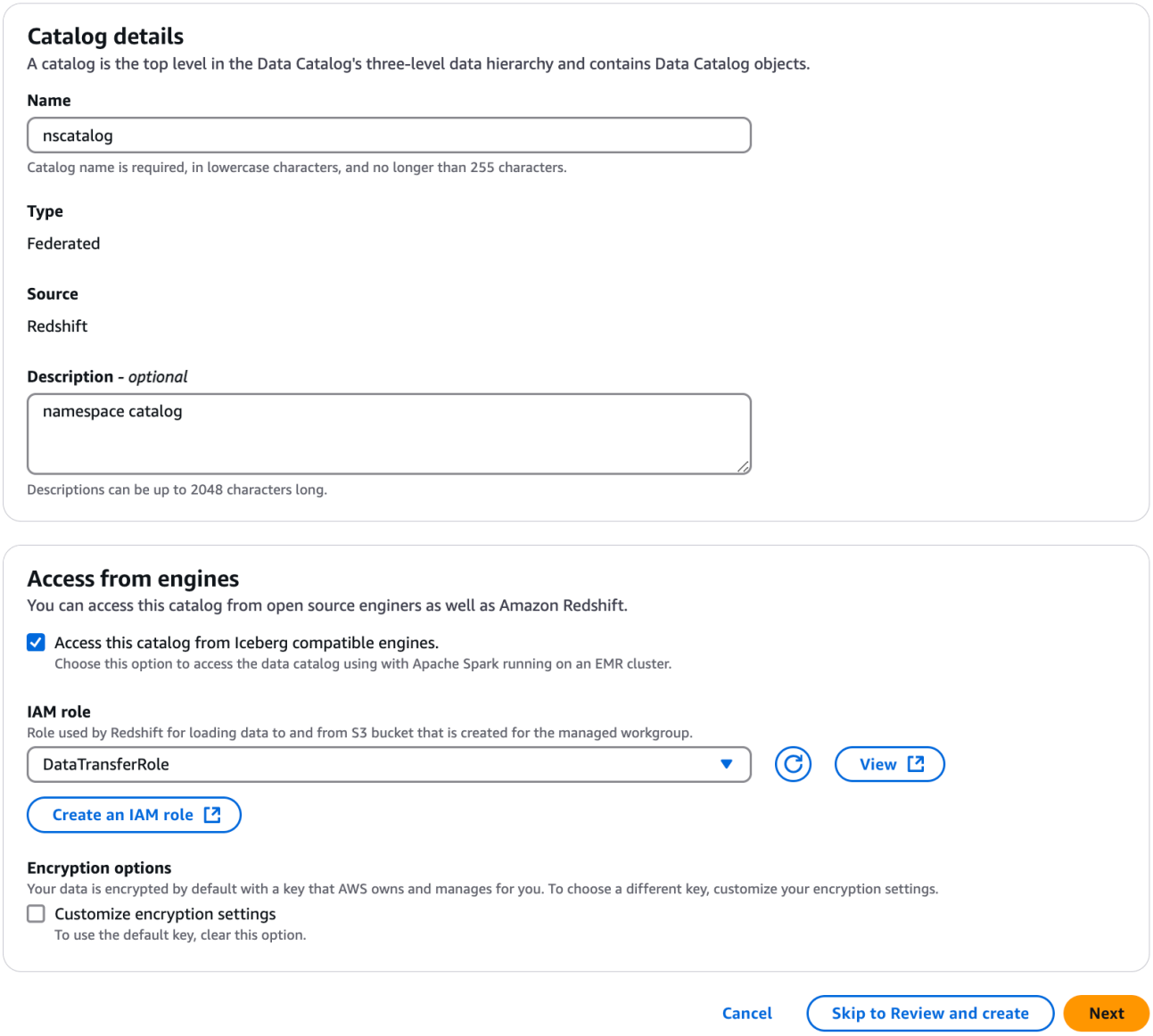
2. 建立聯合目錄。

接受邀請後，您需要在 Data Catalog 中建立聯合目錄，將 Amazon Redshift 命名空間中的物件映射至 Data Catalog。您必須是資料湖管理員或具有建立目錄所需許可的使用者或角色。

Console

1. 接受命名空間邀請後，即會顯示設定目錄詳細資訊頁面。
2. 在設定目錄詳細資訊頁面上，輸入目錄的唯一名稱。針對目錄名稱使用小寫。目錄名稱長度必須小於或等於 255 個字元。您可以在中繼資料階層 (catalogid.dbName.schema.table).
3. 輸入目錄的描述。描述長度必須小於或等於 2048 個字元。
4. 接著，選擇從 Iceberg 相容引擎存取此目錄核取方塊，以使用 Apache Iceberg 相容分析引擎存取 Amazon Redshift 資源，例如 Amazon EMR 上的 Athena 和 Apache Spark。

   您不需要啟用資料湖存取，即可使用 Amazon Redshift 存取聯合目錄。



5. 為了讓這些查詢引擎能夠讀取和寫入 Amazon Redshift 命名空間，會使用執行讀取和寫入操作所需的運算和儲存資源來 AWS Glue 建立受管 Amazon Redshift 叢集，而不會影響 Amazon Redshift 資料倉儲工作負載。

   您也需要為 IAM 角色提供在 Amazon S3 儲存貯體之間傳輸資料所需的許可。

6. 根據預設，Amazon Redshift 叢集中的資料會使用 AWS 受管金鑰加密。Lake Formation 提供建立自訂 KMS 金鑰以進行加密的選項。如果您使用客戶受管金鑰，則必須將特定金鑰政策新增至金鑰。

   如果您使用客戶受管金鑰來加密 Amazon Redshift 叢集/命名空間中的資料，請選擇自訂加密設定。若要使用自訂金鑰，您必須將其他自訂受管金鑰政策新增至 KMS 金鑰。如需詳細資訊，請參閱在 中管理 Amazon Redshift 命名空間的先決條件 AWS Glue Data Catalog。

AWS CLI

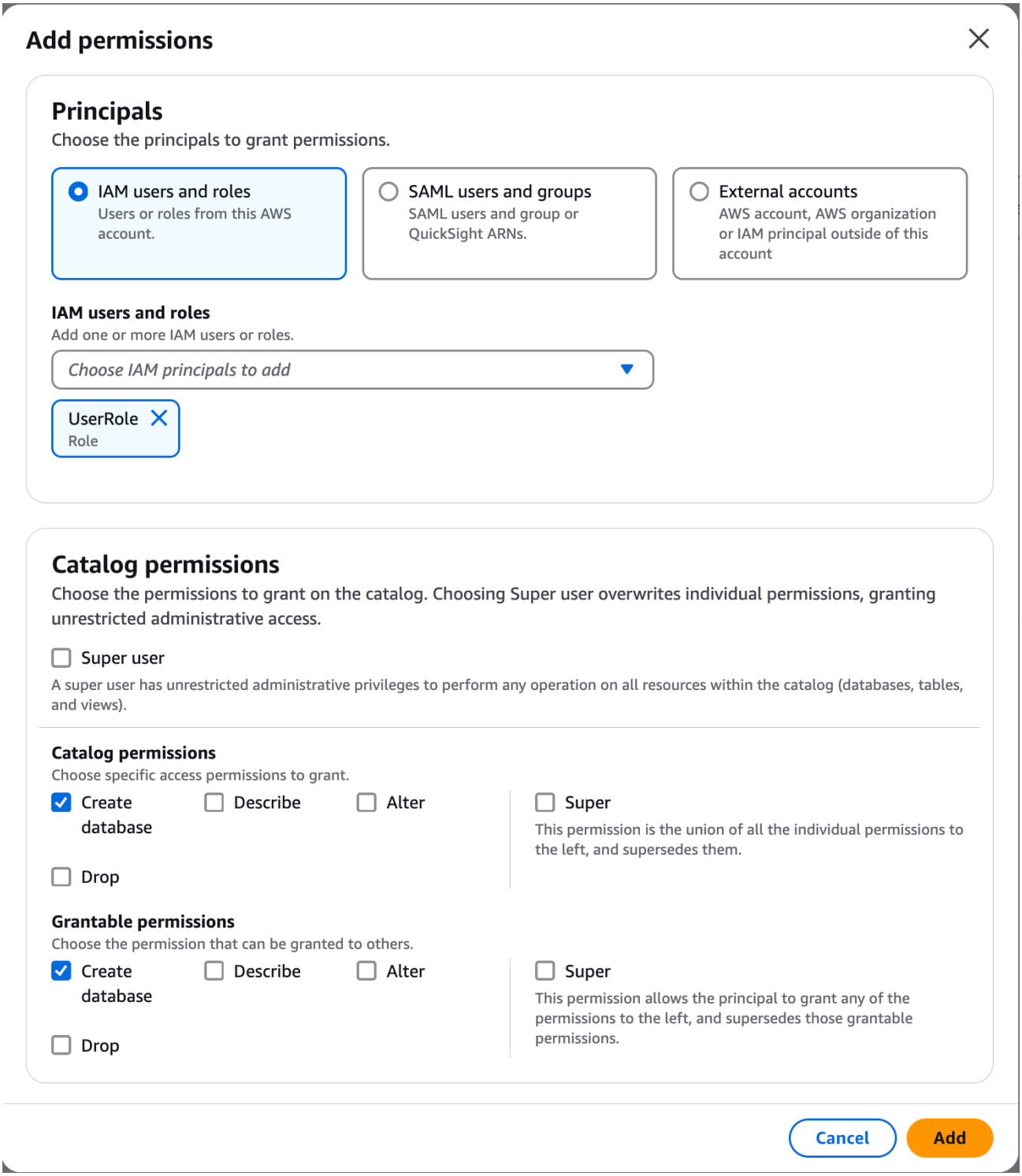
使用下列範例程式碼，使用 發佈至 Data Catalog 的 Amazon Redshift 資料來建立目錄 AWS CLI。

```
aws glue create-catalog
--cli-input-json \
'{
    "Name": "nscatalog",
    "CatalogInput": {
        "Description": "Redshift federated catalog",
        "CreateDatabaseDefaultPermissions" : [],
        "CreateTableDefaultPermissions": [],
        "FederatedCatalog": {
            "Identifier": "arn:aws:redshift:us-
east-1:123456789012:datashare:11524d7f-f56d-45fe-83f7-d7bb0a4d6d71/
ds_internal_namespace",
            "ConnectionName": "aws:redshift"
        },
        "CatalogProperties": {
          "DataLakeAccessProperties" : {
            "DataLakeAccess" : true,
            "DataTransferRole" :
 "arn:aws:iam::123456789012:role/DataTransferRole"
         }
       }
    }
}'
```

3. 將許可授予您帳戶或外部帳戶中的使用者。

AWS Management Console

1. 選擇下一步，將許可授予共用目錄、資料庫和資料表上的其他使用者。
2. 在新增許可畫面上，選擇要授予的委託人和許可類型。



a. 在委託人區段中，選擇委託人類型，然後指定委託人以授予許可。

- IAM 使用者和角色 – 從 IAM 使用者和角色清單中選擇一或多個使用者或角色。

- SAML 使用者和群組 – 對於 SAML 和 Amazon QuickSight 使用者和群組，輸入透過 SAML 聯合的使用者或群組的一或多個 Amazon Resource Name (ARNs)，或 Amazon QuickSight 使用者或群組ARNs。在每個 ARN 之後按 Enter。

  如需如何建構 ARNs 的資訊，請參閱 AWS CLI 授予和撤銷 AWS CLI 命令。

- 外部帳戶 – 針對 AWS、 AWS 組織或 IAM 主體，輸入一或多個有效的 AWS 帳戶 IDs、組織 IDs、組織單位 IDs 或 IAM 使用者或角色的 ARN。在每個 ID 之後按 Enter。組織 ID 包含「o-」，後面接著 10-32 個小寫字母或數字。組織單位 ID 以「ou-」開頭，後面接著 4-32 個小寫字母或數字（包含 OU 的根 ID)。此字串後面接著第二個「-」破折號和 8 到 32 個額外的小寫字母或數字。

b. 在許可區段中，選取許可和可授予的許可。

   在目錄許可下，選取要授予的一或多個許可。在可授予許可下，選取授予收件人可以授予其 AWS 帳戶中其他委託人的許可。當您從外部帳戶將許可授予 IAM 主體時，不支援此選項。

   選擇超級使用者，將無限制許可授予使用者目錄中的資源（資料庫、資料表、檢視）。

3. 選擇新增。

AWS CLI

使用下列範例來使用 授予目錄、資料庫和資料表許可 AWS CLI：

- 下列範例顯示如何授予聯合目錄的許可。

```
aws lakeformation grant-permissions
 --cli-input-cli-json \
   '{
         "Principal": {
              "DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:role/
non-admin"
          },
          "Resource": {
               "Catalog": {
                     "Id": "123456789012:nscatalog"
                }
            },
            "Permissions": [
                   "DESCRIBE","CREATE_CATALOG"
```

```
            ],
           "PermissionsWithGrantOption": [
            ]
    }'
```

- 使用下列範例授予資料庫的許可。

```
aws lakeformation grant-permissions \
  --cli-input-json \
          '{
              "Principal": {

 "DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:role/non-admin"
              },
              "Resource": {
                  "Database": {
                      "CatalogId": "123456789012:nscatalog/dev",
                      "Name": "public"
                  }
              },
              "Permissions": [
                  "ALL"
              ]
          }'
```

- 下列範例顯示如何授予 Amazon Redshift 資料庫中資料表的許可。

```
aws lakeformation grant-permissions \
  --cli-input-json \
        '{
            "Principal": {

 "DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:role/non-admin"
            },
            "Resource": {
                "Table": {
                    "CatalogId": "123456789012:nscatalog2/dev",
                    "DatabaseName": "public",
                    "TableWildcard" : {}
                }
            },
            "Permissions": [
                "ALL"
```

```
            ]
        }'
```

4. 選擇下一步以檢閱目錄詳細資訊並建立聯合目錄。新建立的聯合目錄和目錄物件會出現在目錄頁面中。

   Amazon Redshift 聯合目錄與 一起參考`catalogID = 123456789012:Redshift-federated catalog id`。

檢視目錄物件

建立聯合目錄之後，您可以使用 Lake Formation 主控台或 檢視目錄中的物件 AWS CLI。

AWS Management Console

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。
2. 在資料目錄下選擇目錄。
3. 從目錄頁面上的清單中選擇聯合目錄。
4. 目錄摘要頁面會顯示您擁有許可的目錄物件 （資料庫和資料表）。許可索引標籤顯示已獲得這些物件許可的 IAM 主體。

AWS CLI

- 下列 AWS CLI 範例示範如何請求最上層目錄。

```
aws glue get-catalog \
--catalog-id 123456789012:nscatalog
```

回應

```
{
    "Catalog": {
        "CatalogId": "123456789012:nscatalog",
        "Name": "nscatalog",
        "ResourceArn": "arn:aws:glue:us-east-1:123456789012:catalog/nscatalog",
        "Description": "Redshift published Catalog",
        "CreateTime": "2024-09-05T14:49:16-07:00",
        "FederatedCatalog": {
```

```
            "Identifier": "arn:aws:redshift:us-
east-1:123456789012:datashare:b1234589-e823-4a14-ad8e-077085540a50/
ds_internal_namespace",
            "ConnectionName": "aws:redshift"
        },
        "CatalogProperties": {
            "DataLakeAccessProperties": {
                "DataLakeAccess": true,
                "DataTransferRole": "arn:aws:iam::123456789012:role/
DataTransferRole",
                "KmsKey": "AWS_OWNED_KMS_KEY",
                "ManagedWorkgroupName": "123456789012:nscatalog",
                "ManagedWorkgroupStatus": "AVAILABLE",
                "RedshiftDatabaseName": "dev"
            }
        },
        "CatalogIdentifier": "e2309c2c2fb048f1a3069dfdc1c7883e",
        "CreateTableDefaultPermissions": [],
        "CreateDatabaseDefaultPermissions": []
    }
}
```

- 下列範例顯示如何請求帳戶中的所有目錄。

```
aws glue get-catalogs \
  --recursive
```

- 下列範例請求顯示如何取得 Amazon Redshift 資料庫層級目錄。

```
aws glue get-catlog \
 --catalog-id 123456789012:namespace catalog name/redshift database name
```

- 下列範例請求顯示如何取得 Amazon Redshift 資料庫層級目錄中的資料庫。

```
aws glue get-databases \
--catalog-id 123456789012:namespace catalog name/redshift database name
```

- 下列範例請求顯示如何在目錄中取得 Amazon Redshift 資料表。

```
aws glue get-table \
  --catalog-id 123456789012:parent catalog name/redshift database \
```

```
  --database-name redshift schema name \
  --name table name
```

- 下列範例示範如何從 Amazon Redshift 資料庫取得所有資料表。

```
aws glue get-tables \
 --catalog-id 123456789012:namespace catalog name/redshift database name \
 --database-name RS schema name
```

更新聯合目錄

您可以使用 Lake Formation 主控台、 AWS CLI 或 UpdateCatalog API 操作，在 Data Catalog 中更新 Amazon Redshift 聯合目錄。

AWS Management Console

請依照下列步驟，使用 Lake Formation 主控台更新您的聯合型錄。

1. 登入 AWS Management Console，並在 https：//https://console.aws.amazon.com/lakeformation/ 開啟 Lake Formation 主控台。
2. 在左側導覽窗格中，選擇資料目錄下的目錄。
3. 在目錄頁面上，選擇您要更新的 Amazon Redshift 聯合目錄。
4. 在動作下，選擇編輯。
5. 在設定目錄詳細資訊畫面的從引擎存取區段下，選擇從 Iceberg 相容引擎存取此目錄。勾選此選項將啟用 Apache Iceberg 相容查詢引擎的資料湖存取。
6. 接著，建立新的 IAM 角色，或選擇具有 政策的現有 IAM 角色，該政策授予許可，以執行進出 Amazon S3 儲存貯體的資料傳輸。

   如需許可的詳細資訊，請參閱 在 中管理 Amazon Redshift 命名空間的先決條件 AWS Glue Data Catalog。
7. 根據預設，Amazon Redshift 叢集中的資料會使用 加密 AWS 受管金鑰。如果您選擇使用客戶受管金鑰加密資料，請建立 KMS 金鑰或選擇具有 在 中管理 Amazon Redshift 命名空間的先決條件 AWS Glue Data Catalog區段中定義之許可的現有金鑰。
8. 選擇儲存。

   成功完成時，目錄詳細資訊頁面會顯示狀態為「成功」的受管工作群組名稱。

AWS CLI

以下是透過將 DataLakeAacess 參數值設定為 update-catalog 來停用資料湖存取的 CLI 輸入範例false。

```
aws glue update-catalog  --cli-input-json \
'{
    "Name": "nscatalog",
    "CatalogInput": {
        "Description": "Redshift published catalog",
        "CreateDatabaseDefaultPermissions" : [],
        "CreateTableDefaultPermissions": [],
        "FederatedCatalog": {
            "Identifier": "arn:aws:redshift:us-
east-1:123456789012:datashare:11524d7f-f56d-45fe-83f7-d7bb0a4d6d71/
ds_internal_namespace",
            "ConnectionName": "aws:redshift"
        },
        "CatalogProperties": {
          "DataLakeAccessProperties" : {
            "DataLakeAccess" : false
        }
       }
    }
}'
```

存取共用聯合目錄

AWS Lake Formation 跨帳戶功能可讓使用者安全地跨多個 AWS 帳戶、 AWS 組織共用分散式資料湖，或直接與另一個帳戶中的 IAM 主體共用，以精細存取中繼資料和基礎資料。

Lake Formation 使用 AWS Resource Access Manager (AWS RAM) 服務來促進資源共用。當您與其他帳戶共用目錄資源時， 會 AWS RAM 傳送邀請給承授者帳戶，以接受或拒絕資源授予。

Amazon Athena 和 Redshift Spectrum 等整合式分析服務需要資源連結，才能在查詢中包含共用資源。主體需要在 中建立資源連結 AWS Glue Data Catalog ，以從另一個資源共用資源 AWS 帳戶。如需資源連結的詳細資訊，請參閱資源連結如何在 Lake Formation 中運作。

目錄連結容器是 Data Catalog 物件，參考來自其他 AWS 帳戶的本機或跨帳戶聯合資料庫層級目錄。您也可以在目錄連結容器中建立資料庫連結和資料表連結。當您建立資料庫連結或資料表連結時，您必須指定位於相同目標 Amazon Redshift 資料庫層級目錄 (Amazon Redshift 資料庫) 下的目標資源。

若要建立目錄連結容器，您需要 Lake Formation CREATE_CATALOG或 glue:CreateCatalog許可。

建立跨帳戶聯合目錄的目錄連結容器

您可以使用 AWS Lake Formation 主控台、CreateCatalogAPI 或 AWS Command Line Interface ()， AWS Glue 建立指向任何 AWS 區域中 Redshift 資料庫層級聯合目錄的目錄連結容器AWS CLI。

建立目錄連結容器至共用目錄 （主控台）

1. 在 https：//https://console.aws.amazon.com/lakeformation/ 開啟 AWS Lake Formation 主控台。以具有 Lake Formation CREATE_CATALOG許可的委託人身分登入。
2. 在導覽窗格中，選擇目錄，然後選擇建立目錄。
3. 在設定目錄詳細資訊頁面上，提供下列資訊：

   名稱

   輸入與目錄名稱遵守相同規則的名稱。名稱可以與目標共用目錄相同。

   Type

   選擇目錄連結容器做為目錄類型。

   來源

   選擇 Redshift。

   目標 Redshift 目錄

   選取 Redshift 資料庫層級聯合目錄，或從清單中選擇本機 （擁有） 目錄。

   此清單包含與您的帳戶共用的所有目錄。請注意，目錄擁有者帳戶 ID 會與每個目錄一起列出。如果您沒有看到您知道已與 帳戶共用的目錄，請檢查下列項目：

   - 如果您不是資料湖管理員，請檢查資料湖管理員是否授予目錄上的 Lake Formation 許可。
   - 如果您是資料湖管理員，而且您的帳戶與授予帳戶不在同一個 AWS 組織中，請確定您已接受目錄的 AWS Resource Access Manager (AWS RAM) 資源共享邀請。如需詳細資訊，請參閱從 接受資源共享邀請 AWS RAM。
4. 若要讓 Apache Iceberg 查詢引擎能夠讀取和寫入 Amazon Redshift 命名空間， 會使用執行讀取和寫入操作所需的運算和儲存資源來 AWS Glue 建立受管 Amazon Redshift 叢集，而不會影響 Amazon Redshift 資料倉儲工作負載。您需要為 IAM 角色提供在 Amazon S3 儲存貯體之間傳輸資料所需的許可。
5. 選擇下一步。

6. （選用）選擇新增許可，將許可授予其他委託人。

   不過，授予目錄連結容器的許可並不會授予目標（連結）目錄的許可。您必須分別授予目標目錄的許可，目錄連結才能在 Athena 中顯示。

7. 接著，檢閱目錄連結容器詳細資訊，然後選擇建立目錄。

   然後，您可以在目錄頁面下檢視連結容器名稱。

   現在，您可以在目錄連結容器中建立資料庫連結和資料表連結，以啟用查詢引擎的存取。

建立目錄連結容器 CLI 範例

- 在下列範例中，`TargetRedshiftCatalog` 物件會指定 Amazon Redshift 聯合資料庫層級目錄 (Amazon Redshift 資料庫）的 arn。建立目錄連結容器時，`DataLakeAccess`必須啟用 。

```
aws glue create-catalog \
  --cli-input-json
    '{
        "Name": "linkcontainer",
        "CatalogInput": {
            "TargetRedshiftCatalog": {
               "CatalogArn": "arn:aws:us-east-1:123456789012:catalog/nscatalog/dev"
             },
            "CatalogProperties": {
              "DataLakeAccessProperties" : {
                "DataLakeAccess" : true,
                "DataTransferRole" : "arn:aws:iam::111122223333:role/
DataTransferRole"
             }
           }
        }
    }'
```

在目錄連結容器下建立資源連結

您可以在目錄連結容器下建立資料庫和資料表連結的資源連結。當您建立資料庫資源連結或資料表資源連結時，您必須指定位於連結容器指向的相同目標 Amazon Redshift 資料庫層級目錄 (Amazon Redshift 資料庫）下的目標資源。

您可以使用 AWS Lake Formation 主控台、API 或 AWS Command Line Interface () 建立共用 Amazon Redshift 資料庫或資料表的資源連結AWS CLI。

- 如需詳細說明，請參閱 建立共用 Data Catalog 資料庫的資源連結。

  以下是在目錄連結容器下建立資料庫資源連結 AWS CLI 的範例。

```
aws glue create-database \
  --cli-input-json \
    '{
        "CatalogId": "111122223333:linkcontainer",
        "DatabaseInput": {
            "Name": "dblink",
             "TargetDatabase": {
               "CatalogId": "123456789012:nscatalog/dev",
                "DatabaseName": "schema1"
             }
        }
    }'
```

- 若要在目錄連結容器下建立資料表資源連結，您必須先在本機中建立 AWS Glue 資料庫 AWS Glue Data Catalog ，以包含資料表資源連結。

  如需建立共用資料表的資源連結的詳細資訊，請參閱 建立共用 Data Catalog 資料表的資源連結。

  - 建立資料庫以包含資料表資源連結範例

```
aws glue create-database \
  --cli-input-json \
      '{
          "CatalogId": "111122223333:linkcontainer",
          "DatabaseInput": {
              "Name": "db1",
              "Description": "creating parent database for table link"
          }
      }'
```

  - 建立資料表資源連結範例

```
aws glue create-table \
  --cli-input-json \
```

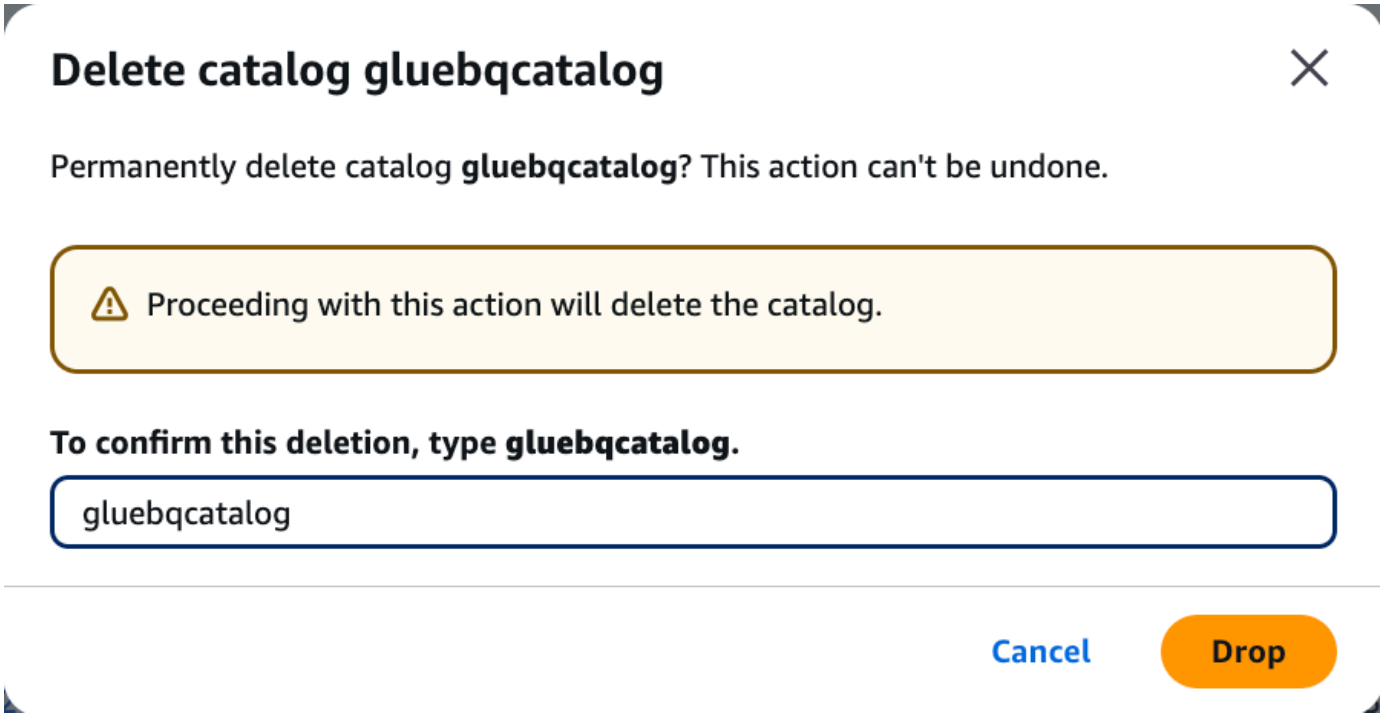
```
'{
    "CatalogId": "111122223333:linkcontainer",
     "DatabaseName": "db1",
    "TableInput": {
        "Name": "tablelink",
        "TargetTable": {
            "CatalogId": "123456789012:nscatalog/dev",
           "DatabaseName": "schema1",
            "Name": "table1"
         }
    }
}'
```

刪除聯合目錄

您可以使用 AWS Glue Data Catalog `glue:DeleteCatalog`操作或 AWS Lake Formation 主控台，刪除您在 中建立的聯合目錄。

刪除聯合目錄（主控台）

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。
2. 在導覽窗格中，選擇資料目錄下的目錄。
3. 從目錄清單中選擇要刪除的目錄。
4. 選擇從動作刪除。
5. 選擇捨棄以確認，聯合目錄將從資料目錄中刪除。



刪除聯合目錄 (CLI)

- ```
  aws glue delete-catalog
   --catalog-id 123456789012:catalog name
  ```

查詢聯合目錄

將許可授予其他主體之後，他們可以使用 Amazon Redshift、Amazon EMR Amazon Athena、 和 AWS Glue ETL 登入 SQL 工具，登入並開始查詢聯合目錄中的資料表。

如需 AWS Glue Data Catalog 使用 Apache Iceberg Rest 延伸端點或獨立 Spark 應用程式連線至 的詳細資訊，請參閱《 開發人員指南》中的存取 AWS Glue Data Catalog一節。 AWS Glue

您可以使用資料定義語言 (DDL) 查詢，在 Amazon EMR 上使用 Apache Spark 在資料庫中建立和管理資料表。若要在 Amazon Redshift 資料庫中建立和刪除資料表，委託人必須具有 Lake Formation `Create table`、 `Drop`許可。

如需授予 Data Catalog 許可的詳細資訊，請參閱 授予 Data Catalog 資源的許可。

如需從 查詢目錄資源的詳細資訊 Amazon Athena，請參閱《Amazon Athena 使用者指南》中的AWS Glue Data Catalog 從 查詢 Amazon Athena。
  ```

其他資源

您可以使用 Amazon SageMaker Lakehouse 來統一存取資料倉儲和資料湖中的資料。透過 SageMaker Lakehouse，您可以透過開放的 Apache Iceberg REST API 使用偏好的分析、機器學習和商業智慧引擎，以協助確保以一致、精細的存取控制安全地存取資料。

- Amazon SageMaker 研討會
- 使用 Amazon SageMaker Lakehouse 簡化企業的資料存取

在 中聯合到外部資料來源 AWS Glue Data Catalog

您可以將 AWS Glue Data Catalog (Data Catalog) 連接到資料倉儲，例如 Amazon Redshift、Snowflake、Amazon RDS 等雲端資料庫 Amazon DynamoDB、Oracle 和 Amazon MSK 等串流服務，以及使用 AWS Glue 連線的 Teradata 等內部部署系統。這些連線存放在 中， AWS Glue Data Catalog 並使用 註冊 AWS Lake Formation，可讓您為每個可用的資料來源建立聯合目錄。

聯合目錄是指向外部資料系統中資料庫的頂層容器。它可讓您直接從外部資料系統查詢資料，而無需擷取、轉換和載入 (ETL) 程序。

如需 AWS Glue 連線的詳細資訊，請參閱《 AWS Glue 開發人員指南》中的連線至資料。

資料湖管理員可以使用 Amazon SageMaker Lakehouse 或 建立聯合目錄Amazon Athena。

然後，資料湖管理員可以使用 Lake Formation 授予目錄內物件的精細許可，控制目錄、資料庫、資料表、資料欄、資料列或儲存格等各種層級的存取。資料分析師可以使用 Athena 來探索和查詢目錄化資料來源，Lake Formation 會強制執行定義的存取政策。分析師可以在單一查詢中跨多個來源聯結資料，而不需要個別連線到每個來源。

主題

- 工作流程
- 將 Data Catalog 連線至外部資料來源的先決條件
- 使用 AWS Glue 連線建立聯合目錄
- 檢視目錄物件
- 刪除聯合目錄
- 查詢聯合目錄
- 其他資源

工作流程

資料湖管理員或具有必要許可的使用者會完成以下步驟，以將 AWS Glue Data Catalog 連線至外部資料來源。

1. 建立與資料來源的 AWS Glue 連線。當您註冊連線時，用於註冊連線的 IAM 角色必須能夠存取 Lambda 函數和 Amazon S3 溢出儲存貯體位置。
2. 向 Lake Formation 註冊連線。
3. 使用連線至可用資料來源的 AWS Glue 連線，在 Data Catalog 中建立聯合目錄。資料庫、資料表和檢視會自動編目在 Data Catalog 中，並向 Lake Formation 註冊。
4. 使用 Lake Formation 許可，將特定目錄、資料庫和資料表的存取權授予資料分析師。您可以使用 Lake Formation 跨資料湖、倉儲和 OLTP 來源定義精細存取控制政策，以啟用資料列層級和資料欄層級的安全篩選條件。

   然後，資料分析師可以使用 Athena 中的 SQL 查詢透過 Data Catalog 存取所有資料，而無需單獨的連線或資料來源憑證。分析師可以執行聯合 SQL 查詢，從多個來源掃描資料，在沒有複雜資料管道的情況下將資料加入就地。

將 Data Catalog 連線至外部資料來源的先決條件

若要將 AWS Glue Data Catalog 連線至外部資料來源、向 Lake Formation 註冊連線，以及設定聯合目錄，您需要完成下列要求：

> **Note**
>
> 我們建議 Lake Formation 資料湖管理員建立 AWS Glue 連線以連線至外部資料來源，並建立聯合目錄。

1. 建立 IAM 角色。

   - 建立具有必要許可的角色，以部署建立與外部資料來源的連線所需的資源 (Lambda 函數、Amazon S3 溢出儲存貯體、IAM 角色和 AWS Glue 連線）。
   - 建立具有必要最低許可的角色來存取 AWS Glue 連線屬性 (Lambda 函數和 Amazon S3 溢出儲存貯體）。這是您在向 Lake Formation 註冊連線時將包含的角色。

若要使用 Lake Formation 來管理和保護資料湖中的資料，您必須向 Lake Formation 註冊 AWS Glue 連線。藉由這樣做，Lake Formation 可以將登入資料提供給 Amazon Athena 以查詢聯合資料來源。

角色必須具有 Amazon S3 儲存貯體和 Lambda 函數的 `Select`或 `Describe`許可。

- s3:ListBucket
- s3:GetObject
- lambda:InvokeFunction

```
{
      "Version": "2012-10-17",
      "Statement": [
        {
          "Effect": "Allow",
          "Action": [
            "s3:*"
          ],
          "Resource": [
               "s3://"+"Your_Bucket_name"+"Your_Spill_Prefix/*",
               "s3://"+"Your_Bucket_name>"+"Your_Spill_Prefix"
          ]
        },
        {
          "Sid": "lambdainvoke",
          "Effect": "Allow",
          "Action": "lambda:InvokeFunction",
          "Resource": "lambda_function_arn"
        },
        {
          "Sid": "gluepolicy",
          "Effect": "Allow",
          "Action": "glue:*",
          "Resource": "*"
        }
      ]
    }
```

- 將下列信任政策新增至用於註冊連線的 IAM 角色：

```
{
```

```
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": [
                    "lakeformation.amazonaws.com"
              ]
            },
            "Action": "sts:AssumeRole"
        }
    ]
}
```

- 註冊連線的資料湖管理員必須具有角色的 iam:PassRole 許可。

  以下是授予此許可的內嵌政策。以有效的 AWS 帳號取代 *<account-id>*，並以角色名稱取代 *<role-name>*。

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

- 若要在 Data Catalog 中建立聯合目錄，請確認您使用的 IAM 角色是 Lake Formation 資料湖管理員，方法是檢查資料湖設定 (aws lakeformation get-data-lake-settings)。

  如果您不是資料湖管理員，則需要 Lake Formation CREATE_CATALOG許可才能建立目錄。下列範例顯示如何授予建立目錄所需的許可。

```
aws lakeformation grant-permissions \
```

```
--cli-input-json \
        '{
            "Principal": {
             "DataLakePrincipalIdentifier":"arn:aws:iam::123456789012:role/non-
admin"
            },
            "Resource": {
                "Catalog": {
                }
            },
            "Permissions": [
                "CREATE_CATALOG",
                "DESCRIBE"
            ]
        }'
```

2. AWS KMS 如果您使用客戶受管金鑰來加密資料來源中的資料，請將下列金鑰政策新增至金鑰。將帳號取代為有效的 AWS 帳號，並指定角色名稱。根據預設，資料會使用 KMS 金鑰加密。Lake Formation 提供建立自訂 KMS 金鑰以進行加密的選項。如果您使用客戶受管金鑰，則必須將特定金鑰政策新增至金鑰。

   如需管理客戶受管金鑰許可的詳細資訊，請參閱客戶受管金鑰。

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "kms:Encrypt",
        "kms:Decrypt",
        "kms:ReEncrypt*",
        "kms:GenerateDataKey*",
        "kms:DescribeKey"
      ],
      "Resource": "arn:aws:kms:us-east-1:123456789012:key/key-1"
    }
  ]
}
```

使用 AWS Glue 連線建立聯合目錄

若要將 AWS Glue Data Catalog 連接到外部資料來源，您需要使用 AWS Glue 啟用與外部資料來源通訊的連線。您可以使用 AWS Glue 主控台、建立 AWS Glue 連線 API 和 Amazon SageMaker Lakehouse 主控台來建立連線。

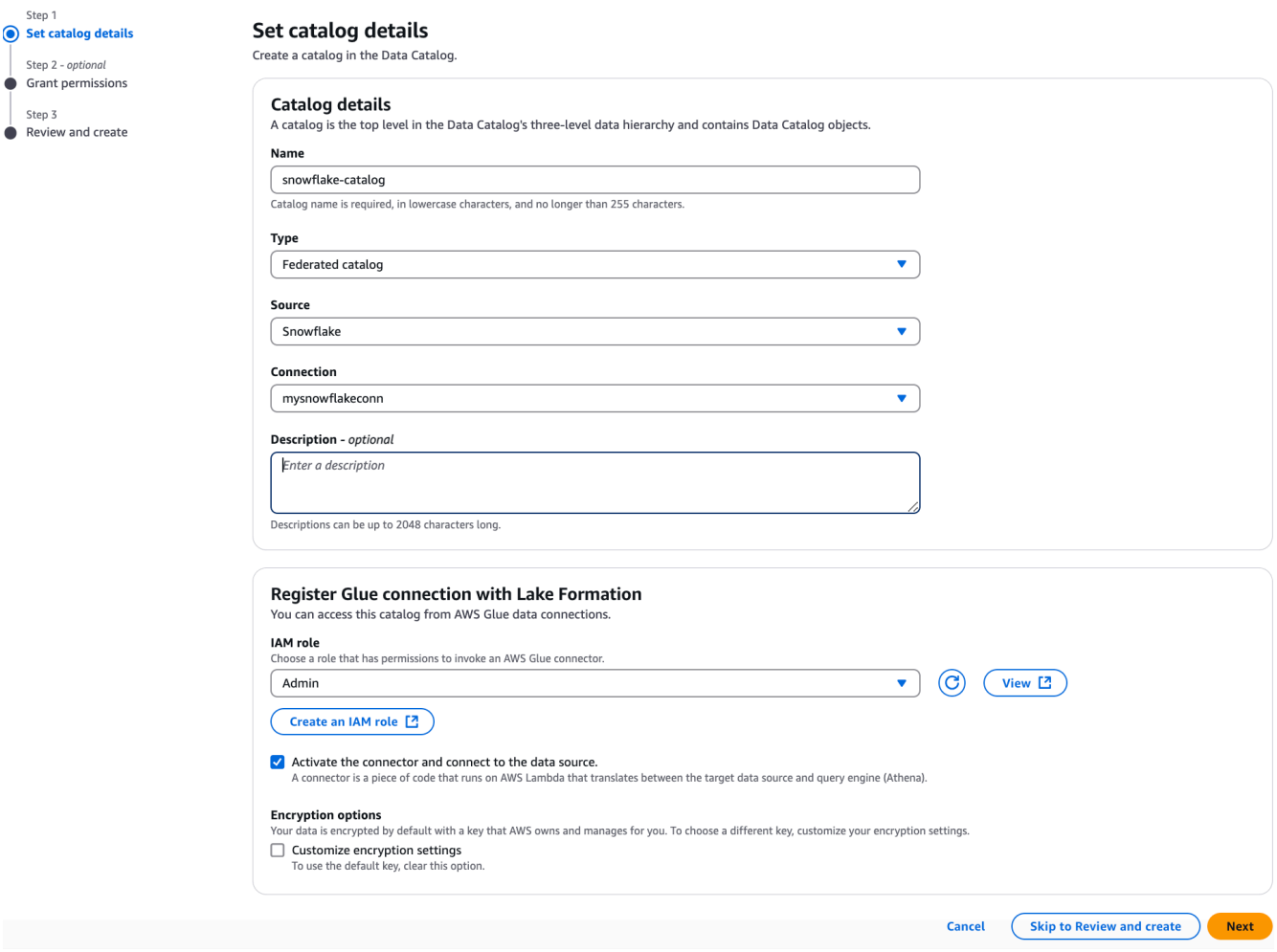
如需建立 AWS Glue 連線的逐步說明，請參閱《 AWS Glue 開發人員指南》中的連線至資料或在 Amazon SageMaker Lakehouse 中建立連線。

當使用者對聯合資料表執行查詢時，Lake Formation 會提供登入資料，以叫用 AWS Glue 連線中指定的 AWS Lambda 函數，從資料來源擷取中繼資料物件。

AWS Management Console

從外部資料來源建立聯合目錄並設定許可 （ 主控台 ）

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。
2. 在導覽窗格中，選擇資料目錄下的目錄。
3. 選取建立目錄選項。
4. 在設定目錄詳細資訊頁面上，輸入下列資訊：



- 名稱 – 聯合目錄的唯一名稱。名稱無法變更，且必須為小寫。名稱最多可包含 255 個字元。帳戶。
- 類型 – 選擇聯合目錄作為目錄類型。
- 來源 – 從下拉式清單中選擇資料來源。隨即顯示您已建立連線的資料來源。如需建立外部資料來源 AWS Glue 連線的詳細資訊，請參閱《 AWS Glue 開發人員指南》中的為連接器建立連線或在 Amazon SageMaker Lakehouse 中建立連線。
- 連線 – 選擇現有的資料來源 AWS Glue 連線。
- 描述 – 輸入從資料來源建立之目錄的描述。

5. 選擇 Lake Formation 的 IAM 角色，以擔任 ，為查詢引擎提供登入資料，以從資料來源存取資料。此角色必須具備必要的許可，才能存取 AWS Glue 連線並叫用 Lambda 函數，才能從外部資料來源存取資料。

   您也可以在 IAM 主控台中建立新的角色。

如需必要的許可，請參閱將 Data Catalog 連線至外部資料來源的先決條件一節。

6. 選取 選項 啟用連接器以連接至資料來源，讓 Athena 能夠執行聯合查詢。

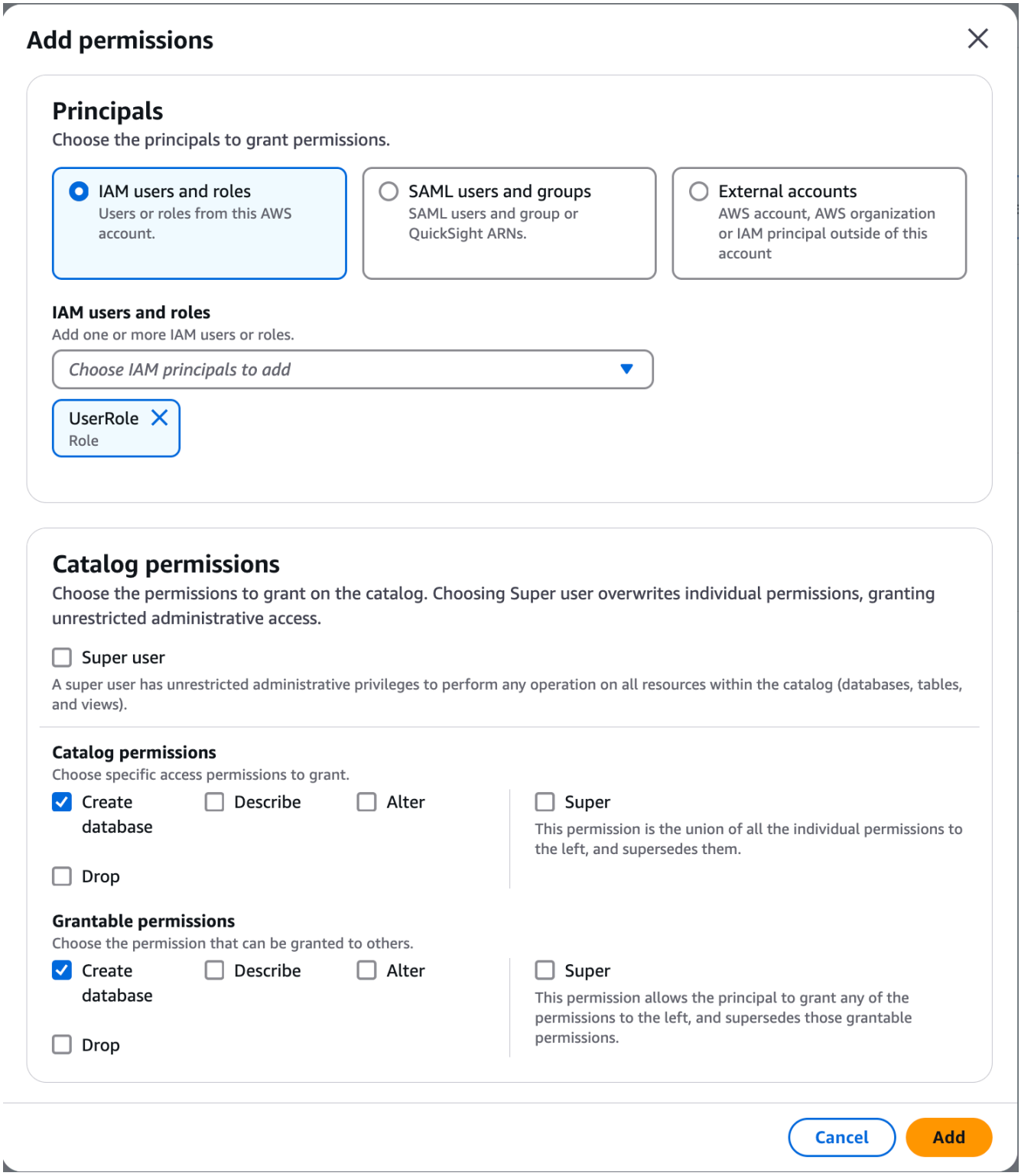
   如需支援的連接器清單，請參閱《Amazon Athena 使用者指南》中的註冊您的連線。

7. 加密選項 – 如果您想要使用自訂金鑰來加密目錄，請選擇自訂加密設定選項。若要使用自訂金鑰，您必須將其他自訂受管金鑰政策新增至 KMS 金鑰。

8. 選擇下一步，將許可授予其他委託人。

9. 在授予許可頁面上，選擇新增許可。

10. 在新增許可畫面上，選擇要授予的委託人和許可類型。



- 在委託人區段中，選擇委託人類型，然後指定委託人以授予許可。
  - IAM 使用者和角色 – 從 IAM 使用者和角色清單中選擇一或多個使用者或角色。

- SAML 使用者和群組 – 對於 SAML 和 Amazon QuickSight 使用者和群組，輸入透過 SAML 聯合的使用者或群組的一或多個 Amazon Resource Name (ARNs)，或 Amazon QuickSight 使用者或群組ARNs。在每個 ARN 之後按 Enter。

- 在許可區段中，選取許可和可授予的許可。

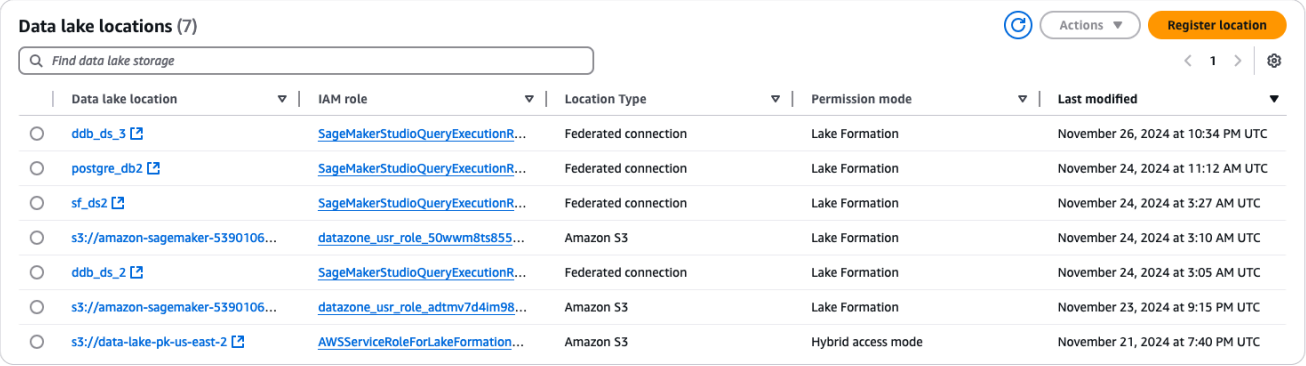
  在目錄許可下，選取要授予的一或多個許可。

  選擇超級使用者，授予目錄內所有資源不受限制的管理許可。

  在授予許可下，選取授予收件人可以授予其 AWS 帳戶中其他委託人的許可。當您從外部帳戶將許可授予 IAM 主體時，不支援此選項。

11. 選擇下一步以檢閱資訊並建立目錄。目錄清單顯示新的聯合目錄。

    資料位置清單會顯示新註冊的聯合連線。



AWS CLI

從外部資料來源建立聯合目錄並設定許可

1. 下列範例示範如何建立 AWS Glue 連線。

```
aws glue create-connection
  --connection-input \
      '{
         "Name": "DynamoDB connection",
         "ConnectionType": "DYNAMODB",
         "Description": "A connection created for DynamoDB",
         "ConnectionProperties": {},
         "AthenaProperties": "spill_prefix": "your_spill_prefix",
         "lambda_function_arn": "Lambda_function_arn",
```

```
          "spill_bucket": "Your_Bucket_name",
          "AuthenticationConfiguration": {}
      }'
```

2. 下列範例示範如何向 Lake Formation 註冊 AWS Glue 連線。

```
aws lakeformation register-resource
  --cli-input-json \
    {"ResourceArn":"arn:aws:glue:us-east-1:123456789012:connection/
dynamo","RoleArn":"arn:aws:iam::123456789012:role/
AdminTelemetry","WithFederation":true}
```

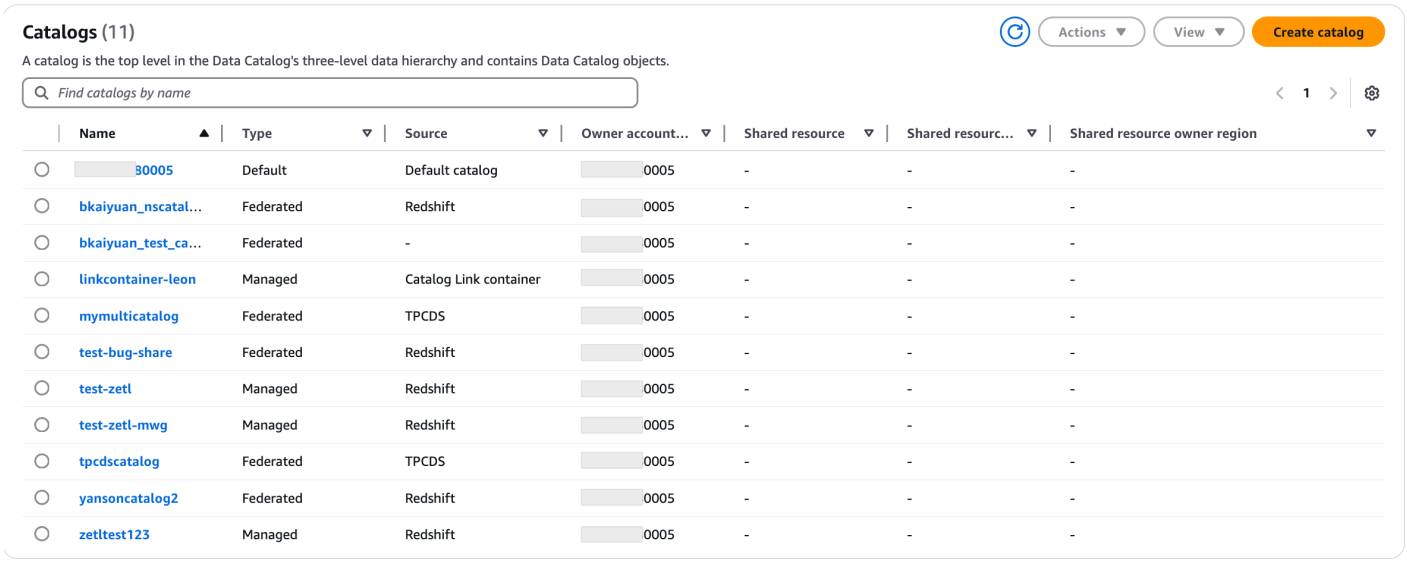
3. 下列範例示範如何建立聯合目錄。

```
aws glue create-catalog
 --cli-input-json \
      '{
       "Name":"ddbcatalog",
       "CatalogInput":{"CatalogProperties":{"DataLakeAccessProperties":
{"DataTransferRole":"arn:aws:iam::123456789012:role/role name"}},
       "CreateDatabaseDefaultPermissions":[],
       "CreateTableDefaultPermissions":[],
       "FederatedCatalog":{"ConnectionName":"dynamo","Identifier":"dynamo"}
         }
       }'
```

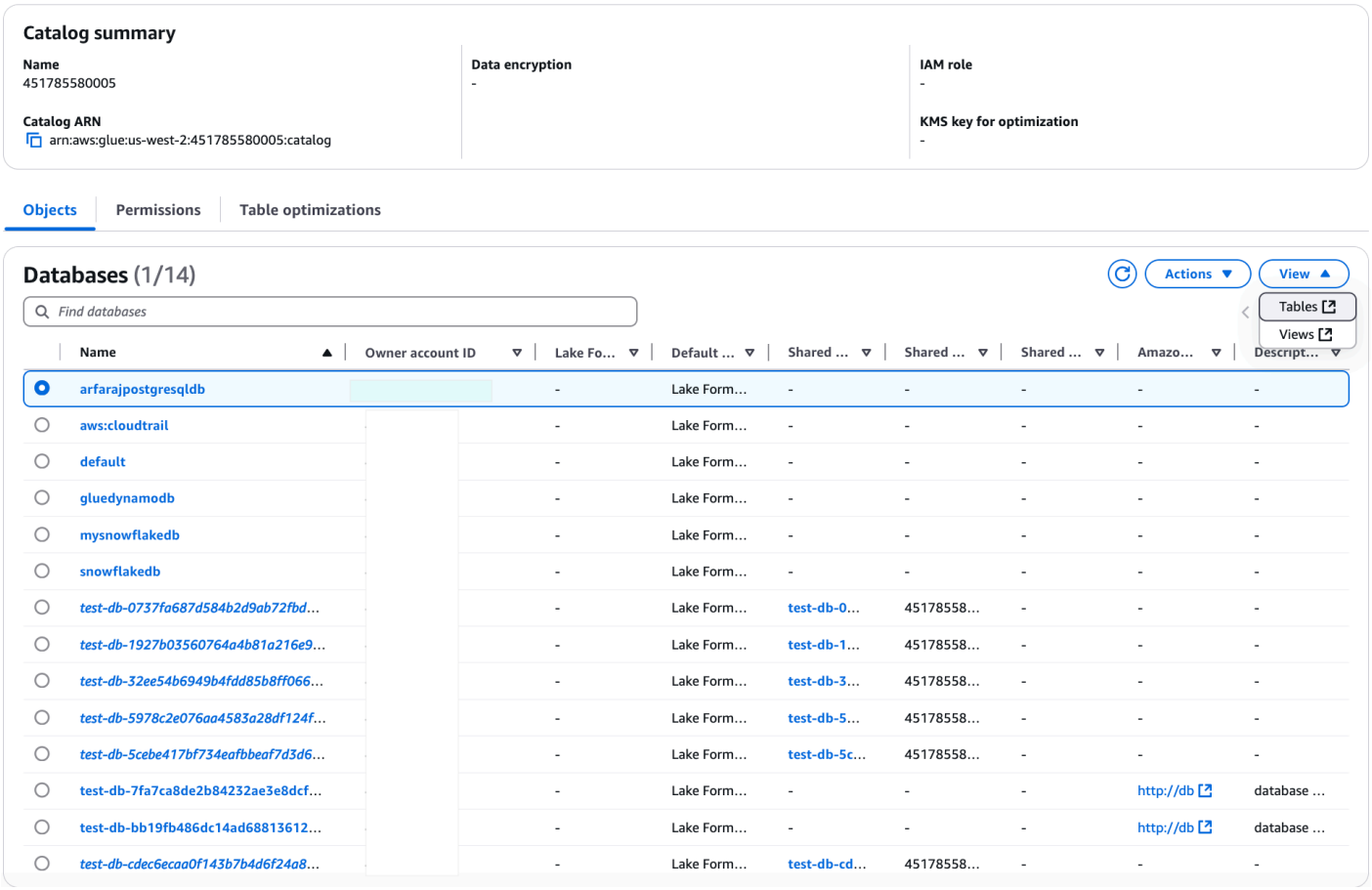
檢視目錄物件

對於每個可用的資料來源， 會在 中 AWS Glue 建立對應的目錄 AWS Glue Data Catalog。建立目錄之後，您可以使用 Lake Formation 主控台或 檢視目錄中的資料庫和資料表 AWS CLI。用於

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。
2. 在資料目錄下選擇目錄。目錄頁面會顯示您擁有許可的目錄。



3. 從清單中選擇目錄，以檢視目錄中包含的資料庫和資料表。此清單包含您帳戶中的資料庫和資源連結，這些是外部帳戶中共用資料庫和資料表的連結，用於跨帳戶存取資料湖中的資料。



4. 選擇檢視下的資料表選項，以檢視和管理資料庫中的資料表。

AWS CLI 檢視目錄和資料庫的範例

下列範例示範如何使用 來檢視目錄 AWS CLI

```
aws glue get-catalog \
--catalog-id 123456789012:dynamodbcatalog
```

下列範例顯示如何請求帳戶中的所有目錄。

```
aws glue get-catalogs \
 --recursive
```

下列範例請求顯示如何取得 目錄中的資料庫。

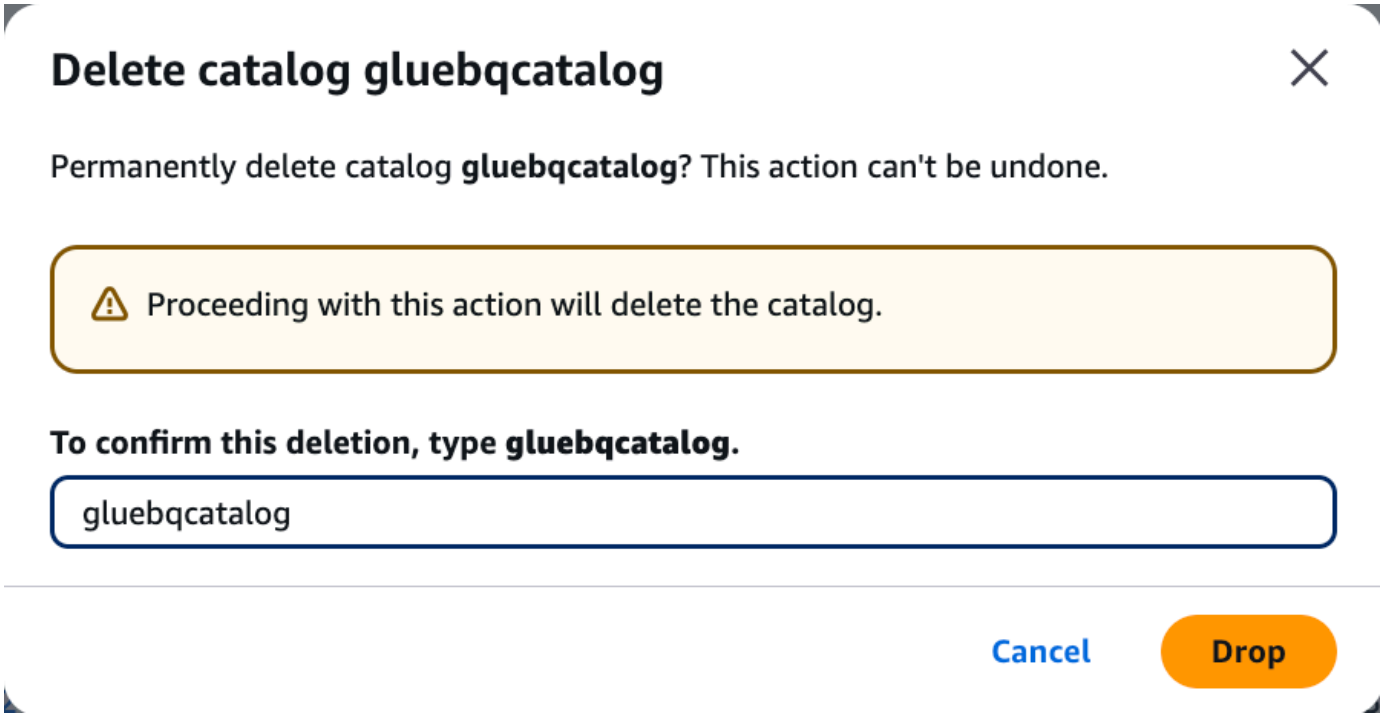
```
aws glue get-database \
--catalog-id 123456789012:dynamodbcatalog
--database-name database name
```

刪除聯合目錄

您可以使用 AWS Glue Data Catalog `glue:DeleteCatalog`操作或 AWS Lake Formation 主控台，刪除您在 中建立的聯合目錄。

刪除聯合目錄 （主控台）

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。
2. 在導覽窗格中，選擇資料目錄下的目錄。
3. 從目錄清單中選擇要刪除的目錄。
4. 選擇從動作刪除。
5. 選擇捨棄以確認，聯合目錄將從資料目錄中刪除。



刪除聯合目錄 (CLI)

- ```
  aws glue delete-catalog
    --catalog-id 123456789012:catalog name
  ```

查詢聯合目錄

將許可授予其他主體之後，他們可以使用 Athena 登入並開始查詢聯合目錄中的資料表。

若要在聯合資料庫中建立和刪除資料表，委託人必須具有 Lake Formation `Create table`的 `Drop` 許可。

如需授予 Data Catalog 許可的詳細資訊，請參閱 授予 Data Catalog 資源的許可。

如需從 查詢資料目錄的詳細資訊 Amazon Athena，請參閱《Amazon Athena 使用者指南》中的AWS Glue Data Catalog 從 查詢 Amazon Athena。

其他資源

在本部落格文章中，我們展示資料分析師如何透過單一、統一的體驗，安全地存取和查詢存放在 S3 資料湖外部的資料，包括 Amazon Redshift 資料倉儲和 Amazon DynamoDB 資料庫。管理員現在可以以
  ```

不同層級的精細程度套用存取控制，以確保敏感資料在擴展資料存取時仍受到保護。這可讓組織加速資料計畫，同時維護安全性和合規性，進而加快資料驅動的決策速度。

- 使用 Amazon SageMaker Lakehouse 編目和管理 Amazon Athena 聯合查詢 Amazon SageMaker

在 中建立 Amazon S3 Tables 目錄 AWS Glue Data Catalog

Amazon S3 Tables 提供專門針對分析工作負載最佳化的 S3 儲存體，可改善查詢效能，同時降低成本。S3 Tables 中的資料會存放在新的儲存貯體類型中：資料表儲存貯體，其會將資料表儲存為子資源。S3 資料表內建支援 Apache Iceberg 標準，可讓您使用常見的查詢引擎，例如 Apache Spark，輕鬆查詢 Amazon S3 資料表儲存貯體中的表格式資料。

您可以將 Amazon S3 資料表儲存貯體和資料表與 AWS Glue Data Catalog （資料目錄） 整合，並從 Lake Formation 主控台或使用服務 APIs 將目錄註冊為 Lake Formation 資料位置。

如需詳細資訊，請參閱《Amazon Simple Storage Service 使用者指南》中的搭配使用 Amazon S3 Tables 與 AWS 分析服務。

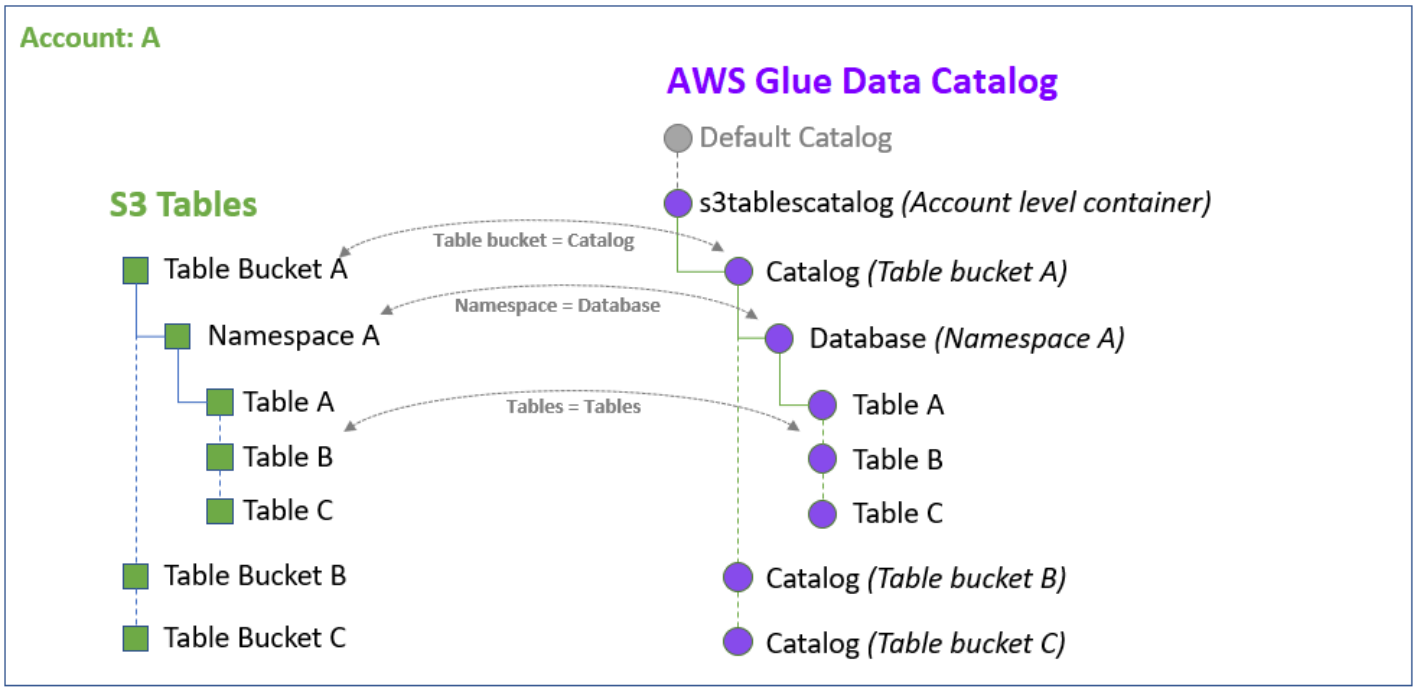
主題

- Data Catalog 和 Lake Formation 整合的運作方式
- 將 Amazon S3 資料表目錄與 Data Catalog 和 Lake Formation 整合的先決條件
- 啟用 Amazon S3 Tables 整合
- 在 S3 資料表目錄中建立資料庫和資料表
- 在另一個 AWS 帳戶中註冊 Amazon S3 資料表儲存貯體
- 授予許可

Data Catalog 和 Lake Formation 整合的運作方式

當您將 S3 資料表目錄與 Data Catalog 和 Lake Formation 整合時， AWS Glue 服務會在您帳戶的預設 Data Catalog `s3tablescatalog`中建立名為 的單一聯合目錄 AWS 區域。整合會以下列方式映射您帳戶和聯合目錄 AWS 區域 下的所有 Amazon S3 資料表儲存貯體資源：

- Amazon S3 資料表儲存貯體會成為 Data Catalog 中的多層級目錄。
- 相關聯的 Amazon S3 命名空間會在 Data Catalog 中註冊為資料庫。
- 資料表儲存貯體中的 Amazon S3 資料表會成為資料目錄中的資料表。



與 Lake Formation 整合後，您可以在資料表儲存貯體目錄中建立 Apache Iceberg 資料表，並透過整合的 AWS 分析引擎 Amazon Athena，例如 Amazon EMR 和第三方分析引擎來存取它們。

將 Amazon S3 資料表目錄與 Data Catalog 和 Lake Formation 整合的先決條件

以下是啟用 Amazon S3 資料表與 AWS Glue Data Catalog 和 整合的先決條件 AWS Lake Formation。

1. AWS 分析服務整合程序已更新。如果您已設定與預覽版本的整合，您可以繼續使用目前的整合。不過，更新的整合程序可提供效能改善。若要更新整合：

   1. 首先，刪除 Lake Formation 中現有的 S3 資料表目錄。若要刪除目錄，請從`S3tablescatalog`目錄清單中選取目錄，然後選擇從動作中刪除。
   2. 接著，取消註冊 的資料位置`S3tablescatalog`。
      a. 在 Lake Formation 主控台的管理區段下，選擇資料位置。
      b. 選取位置，然後從動作功能表中選擇移除。
      c. 出現確認提示時，請選擇移除。

      如需取消註冊資料位置的詳細說明，請參閱 取消註冊 Amazon S3 位置一節。
      d. 然後，遵循 secton 啟用 Amazon S3 Tables 整合 中的更新整合步驟。

2. 當您啟用 Amazon S3 資料表整合時，Lake Formation 會自動註冊 S3 資料表的位置。若要向 Lake Formation 註冊資料表儲存貯體位置，您需要具有 `lakeformation:RegisterResource`、`lakeformation:RegisterResourceWithPrivilegedAccess`和 `lakeformation:CreateCatalog`許可的 IAM 角色/使用者。當具有這些許可的非管理員使用者註冊目錄位置時，Lake Formation 會自動授予他們該位置的`DATA_LOCATION_ACCESS`許可，允許呼叫委託人在已註冊的資料位置上執行所有支援的 Lake Formation 操作。

3. 當您啟用 S3 資料表整合時，您需要選擇 Lake Formation 的 IAM 角色來提供登入資料，以允許資料存取。為對 S3 資料表儲存貯體的 Lake Formation 資料存取建立 IAM 角色。向 Lake Formation 註冊資料表儲存貯體時使用的 IAM 角色需要下列許可：

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "LakeFormationPermissionsForS3ListTableBucket",
            "Effect": "Allow",
            "Action": [
                "s3tables:ListTableBuckets"
            ],
            "Resource": [
                "*"
            ]
        },
        {
            "Sid": "LakeFormationDataAccessPermissionsForS3TableBucket",
            "Effect": "Allow",
            "Action": [
                "s3tables:CreateTableBucket",
                "s3tables:GetTableBucket",
                "s3tables:CreateNamespace",
                "s3tables:GetNamespace",
                "s3tables:ListNamespaces",
                "s3tables:DeleteNamespace",
                "s3tables:DeleteTableBucket",
                "s3tables:CreateTable",
                "s3tables:DeleteTable",
                "s3tables:GetTable",
                "s3tables:ListTables",
                "s3tables:RenameTable",
                "s3tables:UpdateTableMetadataLocation",
```

```
                "s3tables:GetTableMetadataLocation",
                "s3tables:GetTableData",
                "s3tables:PutTableData"
            ],
            "Resource": [
                "arn:aws:s3tables:us-east-1:123456789012:bucket/*"
            ]
        }
    ]
}
```

如需詳細資訊，請參閱用於註冊位置的角色需求。

4. 將下列信任政策新增至 IAM 角色，以允許 Lake Formation 服務擔任角色，並將臨時憑證提供給整合的分析引擎。

```
{
  "Effect": "Allow",
  "Principal": {
    "Service": "lakeformation.amazonaws.com"
  },
  "Action": [
    "sts:AssumeRole",
    "sts:SetSourceIdentity",
    "sts:SetContext"  # add action to trust relationship when using IAM Identity
 center principals with Lake Formation
  ]
}
```

啟用 Amazon S3 Tables 整合

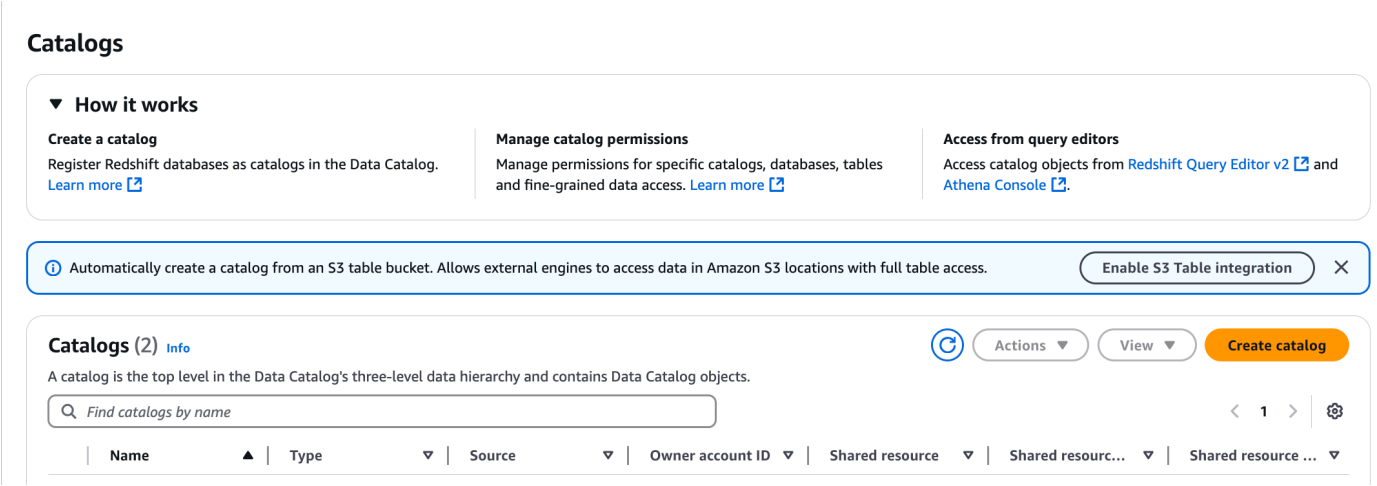
您可以使用 Amazon S3 主控台建立 Amazon S3 資料表儲存貯體，並將其與 AWS 分析服務整合。如需詳細資訊，請參閱搭配 AWS 分析服務使用 Amazon S3 Tables。

在 中 AWS Lake Formation，您可以使用 Lake Formation 主控台或使用 啟用與 AWS Glue Data Catalog 和 的 Amazon S3 Tables AWS Lake Formation 整合 AWS CLI。

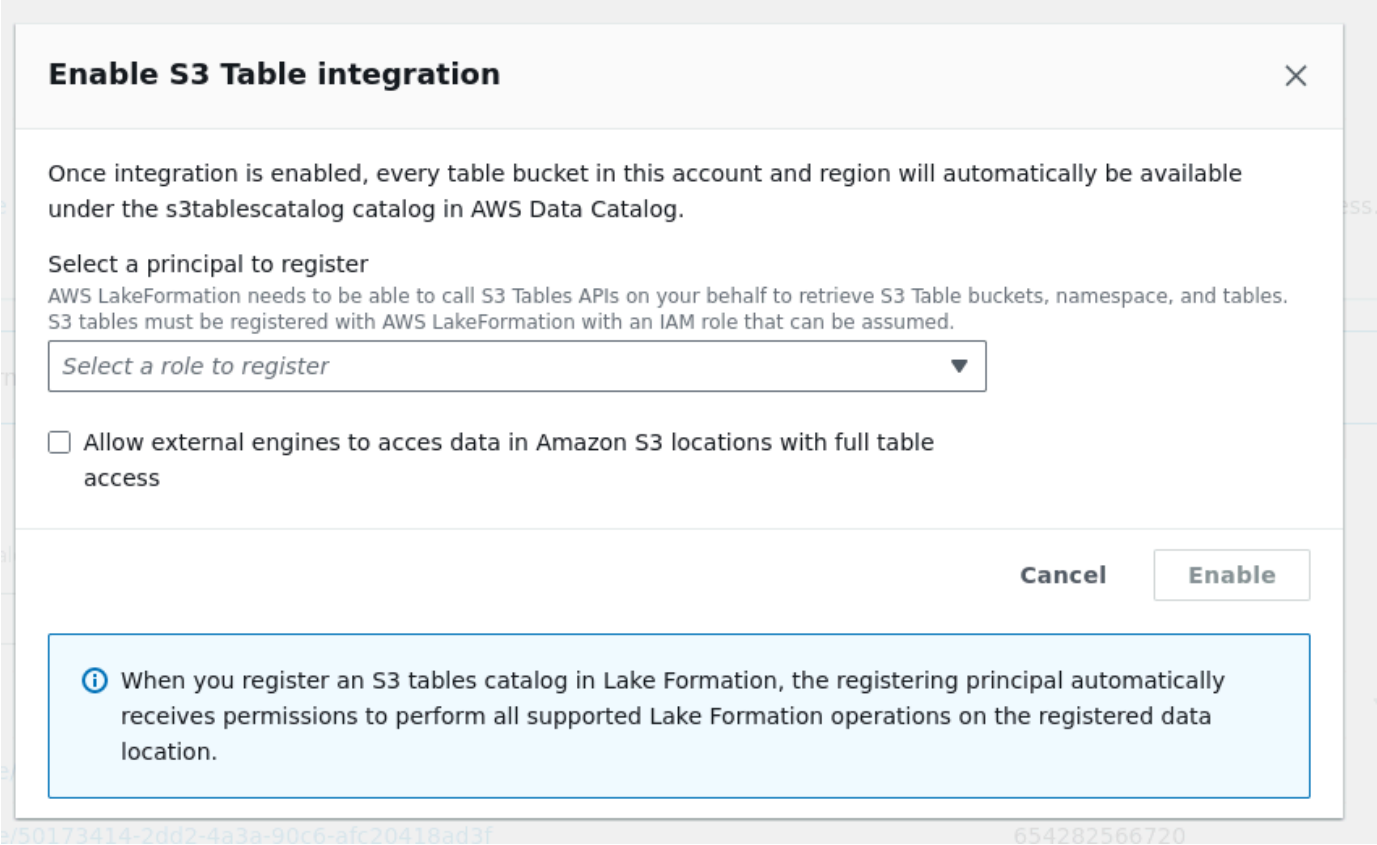
將 Amazon S3 Tables 與 Data Catalog 和 Lake Formation （主控台） 整合

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。
2. 在導覽窗格中，選擇資料目錄下的目錄。

3. 在目錄頁面上選擇啟用 S3 資料表整合。

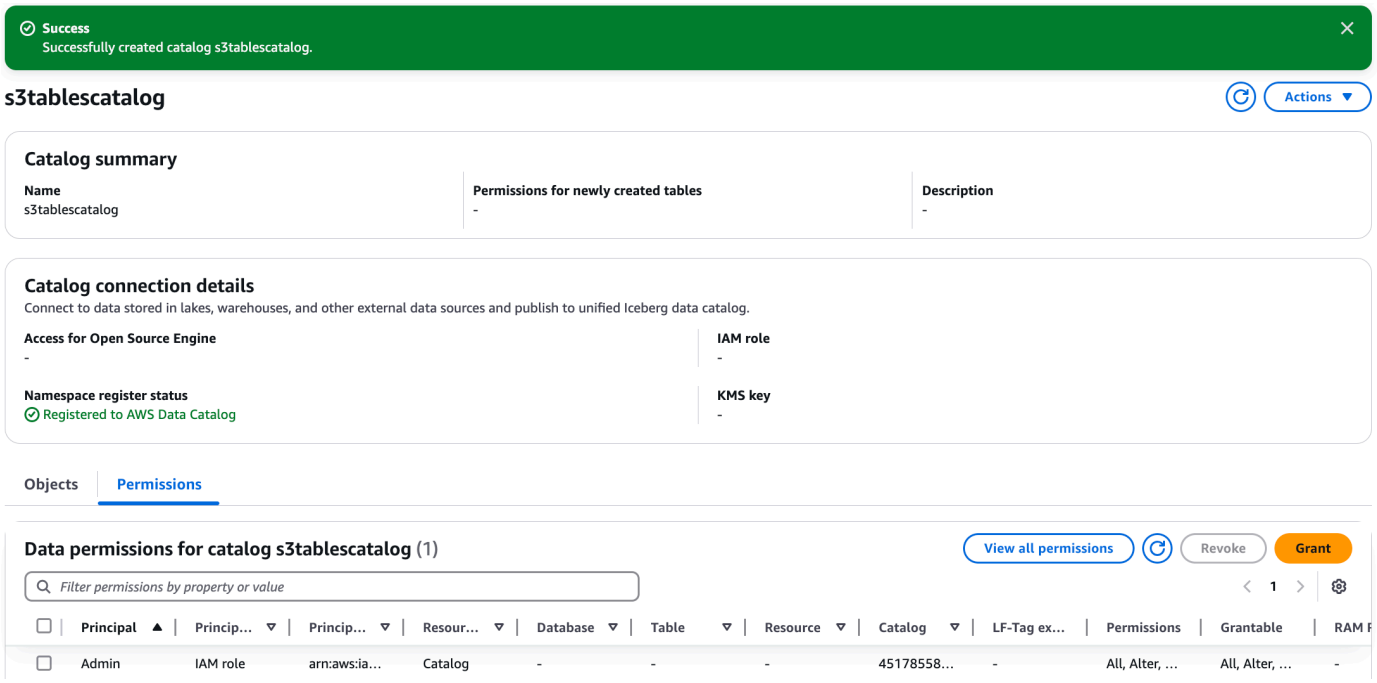


4. 選擇具備 Lake Formation 所需許可的 IAM 角色，以擔任將登入資料提供給分析查詢引擎。如需角色存取資料所需的許可，請參閱先決條件一節step3-permissions中的 。



5. 選取允許外部引擎使用完整資料表存取選項存取 Amazon S3 位置中的資料。當您啟用第三方引擎的完整資料表存取時，Lake Formation 會直接將登入資料傳回至第三方引擎，而不執行 IAM 工作階段標籤驗證。這表示您無法將 Lake Formation 精細存取控制套用至要存取的資料表。

6. 選擇 啟用 。S3 Tables 的新目錄會新增至目錄清單。當您啟用 S3 資料表目錄整合時，服務會向 Lake Formation 註冊 S3 資料表儲存貯體的資料位置。
7. 選擇目錄以檢視目錄物件，並將許可授予其他主體。



若要建立多層級目錄，請參閱《Amazon Simple Storage Service 使用者指南》中的建立資料表儲存貯體一節。

將 Amazon S3 資料表與資料目錄和 Lake Formation (CLI) 整合

1. 將 S3 Tables 目錄註冊為 Lake Formation 資料位置。

```
aws lakeformation register-resource \
  --resource-arn 'arn:aws:s3tables:us-east-1:123456789012:bucket/*' \
  --role-arn 'arn:aws:iam::123456789012:role/LakeFormationDataAccessRole' \
  --with-federation
  --with-privileged-access
```

2. 建立目錄。

```
aws glue create-catalog --cli-input-json file://input.json

'{
   "Name": "s3tablescatalog",
   "CatalogInput" : {
```

```
      "FederatedCatalog": {
         "Identifier": "arn:aws:s3tables:us-east-1:123456789012:bucket/*",
         "ConnectionName": "aws:s3tables"
      },
      "CreateDatabaseDefaultPermissions": [],
      "CreateTableDefaultPermissions": []
   }
}'
```

在 S3 資料表目錄中建立資料庫和資料表

您可以建立資料庫來組織 Apache Iceberg 資料表，以及建立資料表來定義 S3 資料表目錄中資料的結構描述和位置。

建立資料庫 （主控台）

1. 在 https：//https://console.aws.amazon.com/lakeformation/ 開啟 Lake Formation 主控台，並以資料湖管理員或資料庫建立者身分登入。
2. 在導覽窗格中，選擇 Data Catalog 下的資料庫。
3. 選擇建立資料庫。
4. 在建立資料庫頁面上，選擇資料庫選項，然後輸入下列詳細資訊：
   - 名稱 – 資料庫的唯一名稱
   - 資料目錄 – 選擇 S3 資料表目錄。資料庫將位於此目錄中。
   - 描述 –（選用） 新增描述和位置。
   - 新資料表的 IAM 存取控制 – 選擇性地選取僅對此資料庫中的新資料表使用 IAM 存取控制。如需此選項的相關資訊，請參閱變更資料湖的預設設定一節。
   - 選擇建立資料庫。您可以查看在 S3 資料表目錄下建立的資料庫。

使用 建立資料庫 AWS CLI

下列 CLI 命令顯示如何在 S3 資料表目錄中建立資料庫。

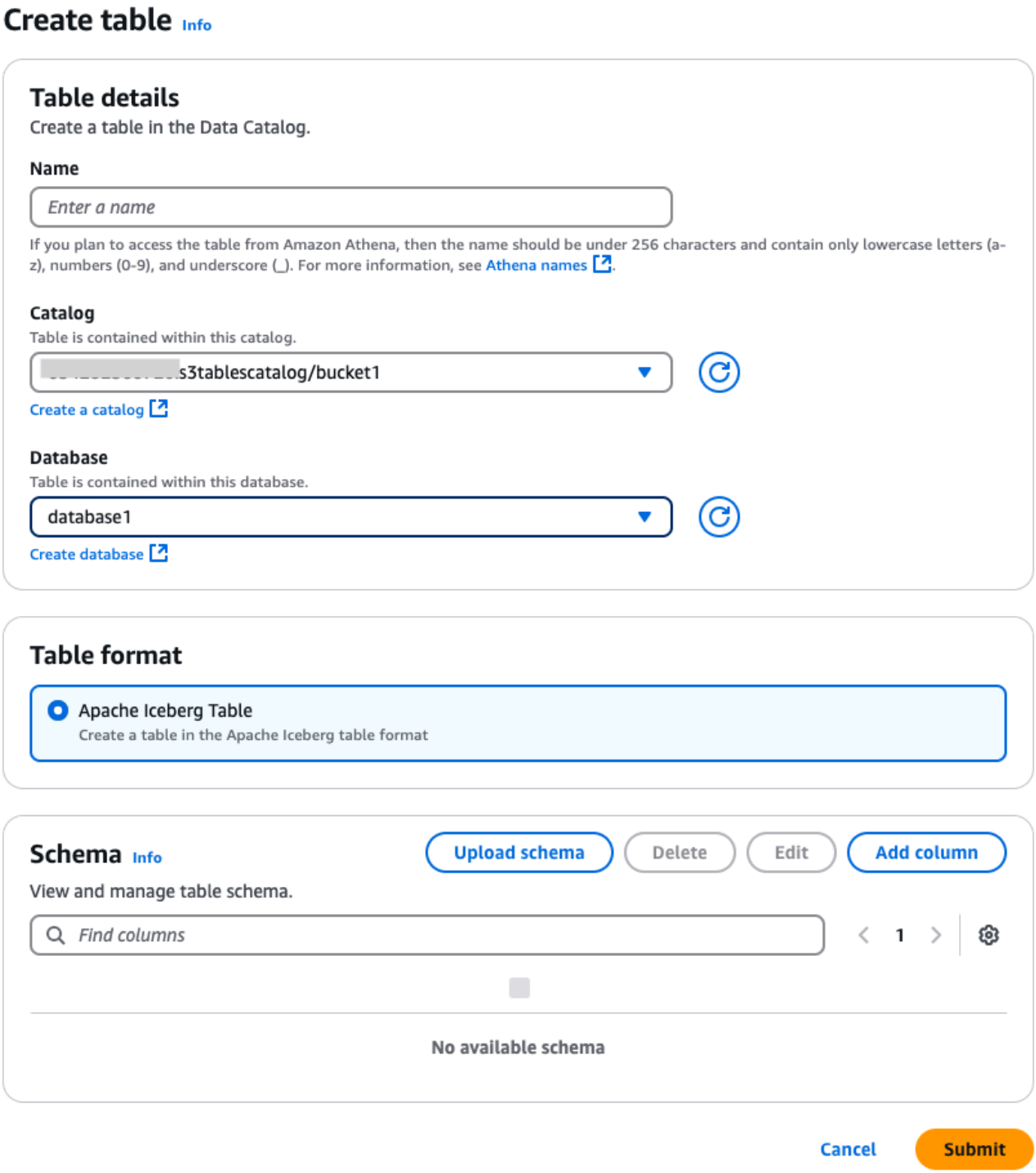
```
aws glue create-database
--region us-east-1 \
--catalog-id "123456789012:s3tablescatalog/test" \
```

```
--database-input \
 '{ "Name": "testglueclidbcreation" }'
```

建立資料表 (AWS Management Console)

您可以使用 Lake Formation 主控台或 API，在 S3 資料表目錄中建立 Apache Iceberg AWS Glue `CreateTable`中繼資料表。

1. 開啟 Lake Formation 主控台，網址為 https：//https://console.aws.amazon.com/lakeformation/`CreateTable`。
2. 在導覽窗格中，選擇 Data Catalog 下的資料表。
3. 選擇建立資料表。
4. 在建立資料表頁面上，輸入資料表詳細資訊：



- 名稱 – 輸入資料表的唯一名稱。
- 目錄 – 選擇 S3 資料表目錄做為目錄。
- 資料庫 – 選擇 S3 資料表目錄下的資料庫。

- 描述 – 輸入資料表的描述。
- 結構描述 – 選擇新增資料欄以新增資料欄和資料欄的資料類型。您可以選擇建立空白資料表，稍後再更新結構描述。Iceberg 可讓您在建立資料表之後發展結構描述和分割區。您可以使用 Athena 查詢來更新資料表結構描述，以及更新分割區的 Spark 查詢。

5. 選擇提交。

建立資料表 (AWS CLI)

```
aws glue create-table \
--database-name "testglueclidbcreation" \
--catalog-id "123456789012:s3tablescatalog/test" \
--region us-east-1 \
--table-input \
'{ "Name": "testtablegluecli", "Parameters": { "format": "ICEBERG" },
 "StorageDescriptor": { "Columns": [ {"Name": "x", "Type": "int", "Parameters":
 {"required": "true"}} ] } }'
```

在另一個 AWS 帳戶中註冊 Amazon S3 資料表儲存貯體

您可以從另一個 AWS 帳戶中的 Lake Formation 帳戶註冊個別 Amazon S3 資料表儲存貯體 (ARN 格式：`arn:aws:s3tables:us-east-1:`*`account-id`*`:bucket/`*`bucket-name`*)。例如，您可以從帳戶 B 的 Lake Formation 中的帳戶 A 註冊資料表儲存貯體。

先決條件

開始跨帳戶儲存貯體註冊之前：

- 在帳戶 A 中建立資料表儲存貯體。
- 在帳戶 B 中建立具有儲存貯體註冊適當許可的 IAM 角色。

  如需向 Lake Formation 註冊資料表儲存貯體所需許可的詳細資訊，請參閱 [將 Amazon S3 資料表目錄與 Data Catalog 和 Lake Formation 整合的先決條件](#) 。
- 在您要建立 S3 資料表目錄的帳戶中註冊資料表儲存貯體。
- 對於跨帳戶存取，註冊資料表儲存貯體時指定的角色必須是具有適當許可的帳戶內角色，才能存取跨帳戶儲存貯體。此角色需要必要的 S3 Tables IAM 動作才能存取儲存貯體資源。

跨帳戶資料表儲存貯體註冊

在下列程序中，帳戶 A 是擁有 帳戶的資源，而帳戶 B 是將註冊資料表儲存貯體以管理存取許可的位置。

1. 登入帳戶 A AWS Management Console 中的 。

   開啟位於 https://console.aws.amazon.com/s3/ 的 Amazon S3 主控台。

2. 建立資料表儲存貯體。如需詳細資訊，請參閱《Amazon S3 使用者指南》中的建立資料表儲存貯體。

3. 在帳戶 B 中註冊資料表儲存貯體。

   使用 從帳戶 A 向帳戶 B 中的 Lake Formation AWS CLI 註冊資料表儲存貯體。

   ```
   aws lakeformation register-resource \
   --resource-arn 'arn:aws:s3tables:us-east-1:account-A-id:bucket/single-bucket-name'
    \
   --role-arn arn:aws:iam::account-B-id:role/role-name \
   --region us-east-1
   ```

   將 account-A-id、single-bucket-name、account-B-id 和 role-name 取代為您的特定值。

4. 接著，為帳戶 B 中的資料表儲存貯體建立目錄。

   使用 建立目錄 AWS CLI。

   ```
   aws glue create-catalog --region us-east-1 \
   --cli-input-json \
   '{
      "Name": "catalog-name",
      "CatalogInput" : {
         "FederatedCatalog": {
            "Identifier": "arn:aws:s3tables:us-east-1:account A:bucket/single-bucket-
   name",
            "ConnectionName": "aws:s3tables"
         },
         "CreateDatabaseDefaultPermissions": [],
         "CreateTableDefaultPermissions": []
      }
   }'
   ```

將 catalog-name、account-A-id 和 bucket-name 取代為您的特定值。

下列 CLI 範例示範如何檢視目錄的詳細資訊。

```
aws glue get-catalog \
  --catalog-id account-id:catalog-name \
  --region us-east-1
```

5. 接著，在帳戶 B 中新建立的目錄中建立資料庫和資料表。

建立資料庫。

```
aws glue create-database \
  --region us-east-1 \
  --catalog-id "account-B-id:catalog-name" \
  --database-input \
'{
  "Name": "database-name"
}'
```

建立 資料表。

```
aws glue create-table \
  --database-name database-name \
  --catalog-id account-B-id:catalog-name\
  --region us-east-1 \
  --table-input \
  '{
        "Name": "table-name",
        "Parameters": {
            "format": "ICEBERG"
        },
        "StorageDescriptor": {
           "Columns": [
        {"Name": "x", "Type": "int", "Parameters": {"required": "true"}}
          ]
        }
}'
```

將 database-name、account-B-id、 catalog-name 和 table-name 取代為您的特定值。

6. 下列範例示範如何檢視 目錄中的物件。

   檢視資料庫詳細資訊。

```
aws glue get-database \
  --name database-name \
  --catalog-id account-B-id:catalog-name \
  --region us-east-1
```

   檢視資料表詳細資訊。

```
aws glue get-table \
  --name table-name \
  --database-name database-name \
  --catalog-id account-B-id:catalog-name \
  --region us-east-1
```

   將 database-name、account-B-id、 catalog-name 和 table-name 取代為您的特定值。

授予許可

將 S3 資料表與 整合之後 AWS Lake Formation，您可以將 S3 資料表目錄和目錄物件 （資料表儲存貯體、資料庫、資料表） 的許可授予帳戶中的其他 IAM 角色和使用者。Lake Formation 許可可讓您為 Amazon Redshift Spectrum 和 Athena 等整合分析引擎的使用者定義資料表、資料欄和資料列層級精細度的存取控制。

您可以透過將 Lake Formation 許可授予外部 AWS 帳戶，與外部帳戶共用資料庫和資料表。然後，使用者可以執行查詢和任務，以跨多個帳戶聯結和查詢資料表。當您與其他帳戶共用目錄資源時，該帳戶中的主體可以操作該資源，就像資源在其資料目錄中一樣。

當您與外部帳戶共用資料庫和資料表時，無法使用超級使用者許可。

如需授予許可的詳細說明，請參閱 管理 Lake Formation 許可一節。

存取共用的 Amazon S3 資料表

在 S3 資料表目錄中授予資料庫或資料表的跨帳戶許可後，若要存取資源，您需要建立共用資料庫和資料表的資源連結。

1. 在目的地帳戶（接收共用資源的帳戶）中，建立資料庫資源連結。如需詳細說明，請參閱 建立共用 Data Catalog 資料庫的資源連結。

   建立資料庫資源連結的 CLI 範例

```
aws glue create-database
--region us-east-1
--catalog-id "111122223333"
--database-input \
'{
  "Name": "s3table_resourcelink",
  "TargetDatabase": {
    "CatalogId": "011426214932:s3tablescatalog/chmni-s3-table-bucket-011426214932",
    "DatabaseName": "s3_table_ns"
  },
  "CreateTableDefaultPermissions": []
}'
```

2. 在資料表上授予跨帳戶許可。

   跨帳戶許可授予的 CLI 範例

```
aws lakeformation grant-permissions \
--region us-east-1 \
--cli-input-json \
'{
    "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:role/
S3TablesTestExecRole"
    },
    "Resource": {
        "Table": {
            "CatalogId": "011426214932:s3tablescatalog/chmni-s3-table-
bucket-011426214932",
            "DatabaseName": "s3_table_ns",
            "Name": "test_s3_iceberg_table"
        }
    },
    "Permissions": [
        "ALL"
    ]
}'
```

3. 在資源連結上授予 Lake Formation DESCRIBE許可。

   授予 資源連結描述許可的 CLI 範例。

```
aws lakeformation grant-permissions \
    --principal DataLakePrincipalIdentifier=arn:aws:iam::111122223333:role/
S3TablesTestExecRole
    --resource Database='{CatalogId=111122223333;, Name=s3table_resourcelink}' \
    --permissions DESCRIBE
```

在 中建立 Amazon Redshift 受管目錄 AWS Glue Data Catalog

您目前可能沒有可用的 Amazon Redshift 生產者叢集或 Amazon Redshift 資料共用，但想要使用 建立和管理 Amazon Redshift 資料表 AWS Glue Data Catalog。您可以使用 `glue:CreateCatalog` API 或 AWS Lake Formation 主控台， AWS Glue 將目錄類型設定為 `Managed`和 `Catalog source` Redshift，以建立受管目錄。此步驟會執行下列動作：

- 在 Data Catalog 中建立目錄
- 將目錄註冊為 Lake Formation 資料位置
- 建立 Amazon Redshift 受管無伺服器工作群組
- 使用資料共用物件連結 Amazon Redshift 無伺服器工作群組和 Data Catalog

建立受管目錄並設定許可 （主控台）

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。
2. 在導覽窗格中，選擇資料目錄下的目錄。
3. 選取建立目錄選項。
4. 在設定目錄詳細資訊頁面上，輸入下列資訊：
   - 名稱 – 受管目錄的唯一名稱。名稱無法變更，且必須為小寫。名稱最多可包含 255 個字元。 帳戶。
   - 類型 – 選擇 `Managed catalog`做為目錄類型。
   - 儲存 – 選擇`Redshift`儲存。
   - 描述 – 輸入從資料來源建立之目錄的描述。

5. 您可以使用在 Amazon EMR on Amazon EC2 上執行的 Apache Spark 應用程式來存取 中的 Amazon Redshift 資料庫 AWS Glue Data Catalog。

   為了讓 Apache Spark 能夠讀取和寫入 Amazon Redshift 受管儲存， AWS Glue 會建立受管 Amazon Redshift 叢集，其中包含執行讀取和寫入操作所需的運算和儲存資源，而不會影響 Amazon Redshift 資料倉儲工作負載。您也需要為 IAM 角色提供在 Amazon S3 儲存貯體之間傳輸資料所需的許可。如需資料傳輸角色所需的許可，請參閱 在 中管理 Amazon Redshift 命名空間的先決條件 AWS Glue Data Catalog區段中的步驟 5。

6. 根據預設，Amazon Redshift 叢集中的資料會使用 AWS 受管金鑰加密。Lake Formation 提供建立自訂 KMS 金鑰以進行加密的選項。如果您使用客戶受管金鑰，則必須將特定金鑰政策新增至金鑰。

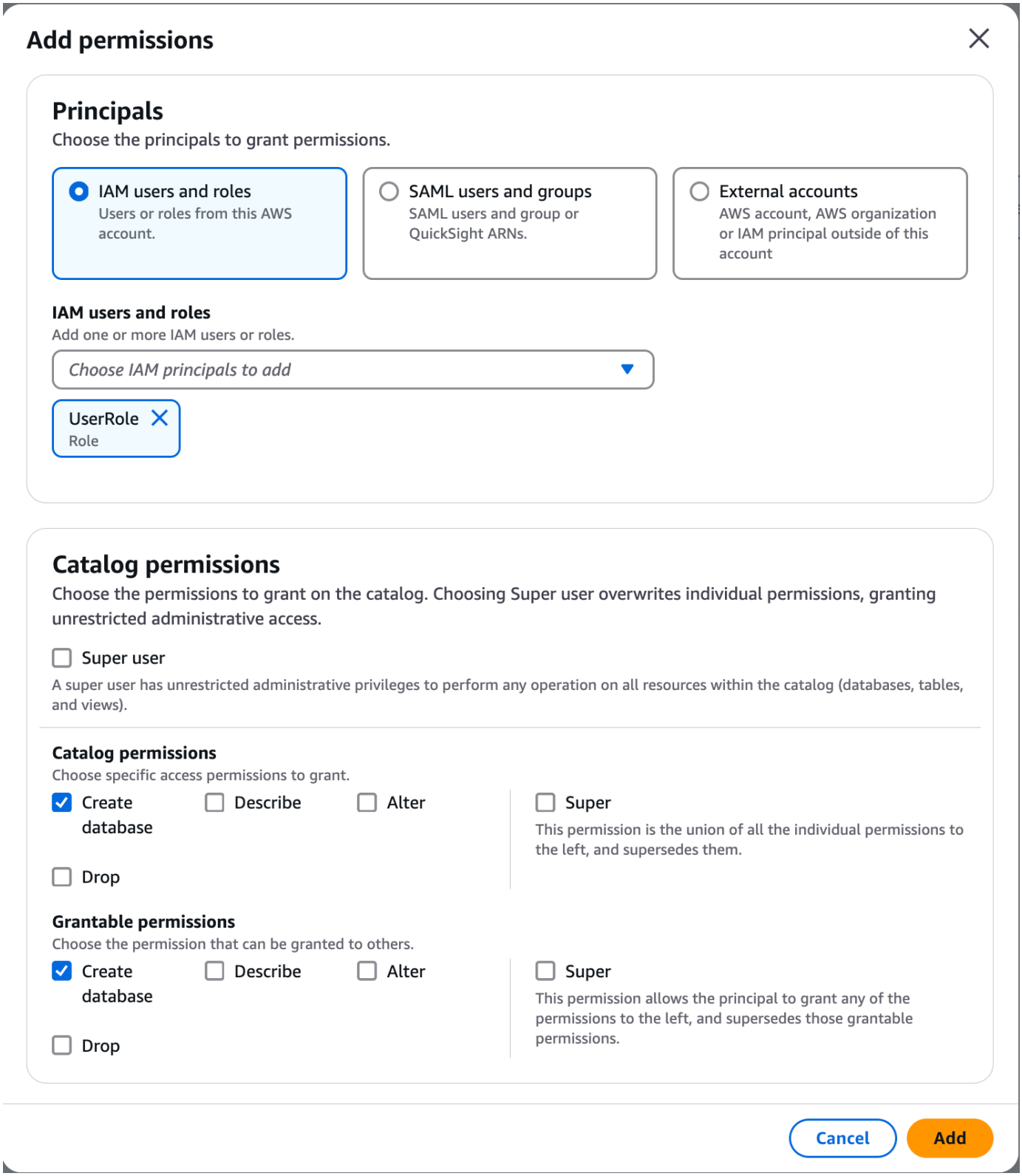
7. 如果您使用客戶受管金鑰來加密 Amazon Redshift 受管儲存中的資料，請選擇自訂加密設定。若要使用自訂金鑰，您必須將其他自訂受管金鑰政策新增至 KMS 金鑰。如需詳細資訊，請參閱在 中管理 Amazon Redshift 命名空間的先決條件 AWS Glue Data Catalog。

8. 加密選項 – 如果您想要使用自訂金鑰來加密目錄，請選擇自訂加密設定選項。若要使用自訂金鑰，您必須將其他自訂受管金鑰政策新增至 KMS 金鑰。

9. 選擇下一步，將許可授予其他委託人。

10. 在授予許可頁面上，選擇新增許可。

11. 在新增許可畫面上，選擇要授予的委託人和許可類型。



- 在委託人區段中，選擇委託人類型，然後指定委託人以授予許可。
  - IAM 使用者和角色 – 從 IAM 使用者和角色清單中選擇一或多個使用者或角色。

- SAML 使用者和群組 – 對於 SAML 和 Amazon QuickSight 使用者和群組，輸入透過 SAML 聯合的使用者或群組的一或多個 Amazon Resource Name (ARNs)，或 Amazon QuickSight 使用者或群組ARNs。在每個 ARN 之後按 Enter。

  如需如何建構 ARNs 的資訊，請參閱 AWS CLI 授予和撤銷 AWS CLI 命令。

- 在許可區段中，選取許可和可授予的許可。

  在目錄許可下，選取要授予的一或多個許可。

  選擇超級使用者，授予目錄內所有資源不受限制的管理許可。

  在授予許可下，選取授予收件人可以授予其 AWS 帳戶中其他委託人的許可。當您從外部帳戶將許可授予 IAM 主體時，不支援此選項。

12. 選擇下一步以檢閱資訊並建立目錄。目錄清單會顯示新的受管目錄。

建立聯合目錄 (CLI)

- 下列範例示範如何建立聯合目錄。

```
aws glue create-catalog --cli-input-json file://input.json

{
    "Name": "CatalogName",
    "CatalogInput": {
      "Description": "Redshift published Catalog",
      "CreateDatabaseDefaultPermissions" : [],
      "CreateTableDefaultPermissions": [],
      "CatalogProperties": {
        "DataLakeAccessProperties" : {
          "DataLakeAccess" : "true",
          "DataTransferRole" : "DTR arn",
          "KMSKey": "kms key arn",  // Optional
          "CatalogType": "aws:redshift"
        }
      }
    }
}
```

Glue get-catalog 回應

```
aws glue get-catalog \
  --catalog-id account-id:catalog-name \
  --region us-east-1

Response:
{
    "Catalog": {
        "Name": "CatalogName",
        "Description": "Glue Catalog for Redshift z-etl use case",
        "CreateDatabaseDefaultPermissions" : [],
        "CreateTableDefaultPermissions": [],
         "CatalogProperties": {
          "DataLakeAccessProperties" : {
            "DataLakeAccess": "true",
            "DataTransferRole": "DTR arn",
            "KMSKey": "kms key arn",
            "ManagedWorkgroupName": "MWG name",
            "ManagedWorkgroupStatus": "MWG status",
            "RedshiftDatabaseName": "RS db name",
            "NamespaceArn": "namespace key arn",
            "CatalogType": "aws:redshift"
          }
       }
    }
```

管理 Amazon Redshift 資料共用中資料的許可

使用 AWS Lake Formation，您可以從 Amazon Redshift 安全地管理資料共用中的資料。Amazon Redshift 是 AWS 雲端中全受管的 PB 級資料倉儲服務。使用資料共用功能，Amazon Redshift 可協助您跨 共用資料 AWS 帳戶。如需 Amazon Redshift 資料共用的詳細資訊，請參閱 Amazon Redshift 中的資料共用概觀。

在 Amazon Redshift 中，生產者叢集管理員會建立資料共用，並與資料湖管理員共用。如需建立資料湖管理員的step-by-step說明，請參閱 建立資料湖管理員。

在您（資料湖管理員）接受資料共用後，您必須為特定資料共用建立 AWS Glue Data Catalog 資料庫。這樣您就可以使用 Lake Formation 許可來控制對它的存取。Lake Formation 會將每個資料共用映射至對應的 Data Catalog 資料庫。這些會在 Data Catalog 中顯示為聯合資料庫。

當資料庫指向 Data Catalog 外部的實體時，稱為聯合資料庫。Amazon Redshift 資料共用中的資料表和檢視會列為資料目錄中的個別資料表。您可以與相同帳戶或 Lake Formation 的其他帳戶中選取的 IAM 主體和 SAML 使用者共用聯合資料庫。您也可以包含資料列和資料欄篩選條件表達式，以限制對特定資料的存取。如需詳細資訊，請參閱Lake Formation 中的資料篩選和儲存格層級安全性。

若要讓使用者存取 Amazon Redshift 資料共用，您必須執行下列動作：

1. 更新 Data Catalog 設定以啟用 Lake Formation 許可。
2. 接受來自 Amazon Redshift 生產者叢集管理員的資料共用邀請，並在 Lake Formation 中註冊資料共用。

   完成此步驟後，您可以在 Lake Formation Data Catalog 中管理資料共用。
3. 建立聯合資料庫並定義該資料庫的許可。
4. 將許可授予資料庫和資料表上的使用者。您可以與相同帳戶或其他帳戶中的使用者共用整個資料庫或資料表的子集。

如需限制的詳細資訊，請參閱Amazon Redshift 資料共用限制。

主題

- 在 Amazon Redshift 資料共用上設定許可的先決條件
- 設定 Amazon Redshift 資料共用的許可
- 查詢聯合資料庫

在 Amazon Redshift 資料共用上設定許可的先決條件

更新預設 Data Catalog 設定

若要啟用 Data Catalog 資源的 Lake Formation 許可，建議您停用 Lake Formation 中的預設 Data Catalog 設定。如需詳細資訊，請參閱變更預設許可模型或使用混合存取模式。

更新許可

除了資料湖管理員許可 (`AWSLakeFormationDataAdmin`) 之外，在 Lake Formation 中接受 Amazon Redshift 資料共用還需要下列許可：

- `glue:PassConnection on aws:redshift`
- `redshift:AssociateDataShareConsumer`

- `redshift:DescribeDataSharesForConsumer`
- `redshift:DescribeDataShares`

資料湖管理員 IAM 使用者隱含具有下列許可。

- data_location_access
- create_database
- lakefomation：registerResource

設定 Amazon Redshift 資料共用的許可

本主題說明您需要遵循的步驟，以接受資料共用邀請、建立聯合資料庫，以及授予許可。您可以使用 Lake Formation 主控台或 AWS Command Line Interface (AWS CLI)。本主題中的範例顯示相同帳戶中的生產者叢集、資料目錄和資料取用者。

若要進一步了解 Lake Formation 跨帳戶功能，請參閱 Lake Formation 中的跨帳戶資料共用。

設定資料共用的許可

1. 檢閱資料共用邀請並接受它。

   Console

   1. 以資料湖管理員身分登入 Lake Formation 主控台，網址為 https：//https://console.aws.amazon.com/lakeformation/。導覽至資料共用頁面。
   2. 檢閱您獲授權存取的資料共用。狀態欄指出您目前的資料共用參與狀態。待定狀態表示您已新增至資料共用，但您尚未接受或拒絕邀請。
   3. 若要回應資料共用邀請，請選取資料共用名稱，然後選擇檢閱邀請。在接受或拒絕資料共用中，檢閱邀請詳細資訊。選擇接受以接受邀請或拒絕以拒絕邀請。如果您拒絕邀請，則無法存取資料共用。

   AWS CLI

   下列範例示範如何檢視、接受和註冊邀請。以有效的 AWS 帳戶 ID 取代 AWS 帳戶 ID。將 取代`data-share-arn`為參考資料共用的實際 Amazon Resource Name (ARN)。

   1. 檢視待定邀請。

```
aws redshift describe-data-shares \
 --data-share-arn 'arn:aws:redshift:us-
east-1:111122223333:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
federatedds' \
```

2. 接受資料共用。

```
 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. 在 Lake Formation 帳戶中註冊資料共用。使用 RegisterResource API 操作在 Lake Formation 中註冊資料共用。 `DataShareArn`是 的輸入參數`ResourceArn`。

> **Note**
>
> 這是必要步驟。

```
aws lakeformation register-resource \
 --resource-arn 'arn:aws:redshift:us-
east-1:111122223333:datashare:abcd1234-1234-ab12-cd34-1a2b3c4d5e6f/
federatedds'
```

2. 建立資料庫。

接受資料共用邀請後，您需要建立指向與資料共用相關聯之 Amazon Redshift 資料庫的資料庫。您必須是資料湖管理員才能建立資料庫。

Console

1. 從邀請窗格中選取資料共用，然後選擇設定資料庫詳細資訊。
2. 在設定資料庫詳細資訊中，輸入資料共用的唯一名稱和識別符。您可以使用此識別符在中繼資料階層 (dbName.schema.table).
3. 選擇下一步，將許可授予共用資料庫和資料表上的其他使用者。

AWS CLI

使用下列範例程式碼建立資料庫，該資料庫指向使用 與 Lake Formation 共用的 Amazon Redshift 資料庫 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. 授予許可。

建立資料庫之後，您可以將許可授予帳戶中的使用者，或外部 AWS 帳戶 和組織。您將無法在映射至 Amazon Redshift 資料共用的聯合資料庫上授予寫入資料許可 （插入、刪除） 和中繼資料許可 （變更、捨棄、建立）。如需授予許可的詳細資訊，請參閱 管理 Lake Formation 許可。

> **Note**
>
> 身為資料湖管理員，您只能檢視聯合資料庫中的資料表。若要執行任何其他動作，您需要對這些資料表授予自己更多許可。

Console

1. 在授予許可畫面上，選取要授予許可的使用者。
2. 選擇 Grant (授予)。

AWS CLI

使用下列範例來使用 授予資料庫和資料表許可 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"
                   ]
```

```
}
```

查詢聯合資料庫

授予許可後，使用者可以使用 Amazon Redshift 登入並開始查詢聯合資料庫。使用者現在可以使用本機資料庫名稱來參考 SQL 查詢中的 Amazon Redshift 資料共用。在 Amazon Redshift 中，透過資料共用共用之公有結構描述`public.customer`中的客戶資料表，將會建立與資料目錄中相同的對應資料表。

1. 使用 Amazon Redshift 查詢聯合資料庫之前，叢集管理員會使用下列命令從 Data Catalog 資料庫建立資料庫：

```
CREATE DATABASE sharedcustomerdb FROM ARN
 'arn:aws:glue:<region>:111122223333:database/tahoedb' WITH DATA CATALOG SCHEMA
 tahoedb
```

2. 叢集管理員會授予資料庫的使用許可。

```
GRANT USAGE ON DATABASE sharedcustomerdb TO IAM:user;
```

3. 您（聯合身分使用者）現在可登入 SQL 工具來查詢資料表。

```
Select * from sharedcustomerdb.public.customer limit 10;
```

如需詳細資訊，請參閱《Amazon Redshift 管理指南》中的查詢 AWS Glue Data Catalog。

管理使用外部中繼存放區的資料集許可

使用 AWS Glue Data Catalog 中繼資料聯合 (Data Catalog 聯合），您可以將 Data Catalog 連線到外部中繼存放區，以存放 Amazon S3 資料的中繼資料，並使用 安全地管理資料存取許可 AWS Lake Formation。您不需要將中繼資料從外部中繼存放區遷移到 Data Catalog。

Data Catalog 提供集中式中繼資料儲存庫，可讓您更輕鬆地管理和探索不同系統中的資料。當您的組織管理 Data Catalog 中的資料時，您可以使用 AWS Lake Formation 來控制對 Amazon S3 中資料集的存取。

> **Note**
>
> 目前，我們僅支援 Apache Hive （第 3 版及更高版本） 中繼存放區聯合。

若要設定 Data Catalog 聯合，我們在 中提供名為 GlueDataCatalogFederation-HiveMetastore 的 AWS Serverless Application Model (AWS SAM) 應用程式 AWS Serverless Application Repository。

參考實作是在 GitHub 上提供，做為AWS Glue Data Catalog 聯合 - Hive 中繼存放區的開放原始碼專案。

AWS SAM 應用程式會建立並部署下列將 Data Catalog 連線至 Hive 中繼存放區所需的資源：

- AWS Lambda 函數 – 託管聯合服務的實作，可在 Data Catalog 和 Hive 中繼存放區之間進行通訊。會 AWS Glue 叫用此 Lambda 函數，以從 Hive 中繼存放區擷取中繼資料物件。
- Amazon API Gateway – Hive 中繼存放區的連線端點，可做為代理，將所有調用路由至 Lambda 函數。
- IAM 角色 – 具有必要許可的角色，可在 Data Catalog 與 Hive 中繼存放區之間建立連線。
- AWS Glue 連線 – 存放 Amazon API Gateway 端點和 IAM 角色以叫用它的一種 AWS Glue 連線 Amazon API Gateway 類型。

當您查詢資料表時， AWS Glue 服務會呼叫 Hive 中繼存放區並擷取中繼資料。Lambda 函數充當 Hive 中繼存放區和資料目錄之間的轉譯器。

建立連線後，若要將 Hive 中繼存放區中的中繼資料與 Data Catalog 同步，您需要使用 Hive 中繼存放區連線詳細資訊在 Data Catalog 中建立聯合資料庫，並將此資料庫映射至 Hive 資料庫。當資料庫指向 Data Catalog 外部的實體時，稱為聯合資料庫。

您可以使用標籤型存取控制和聯合資料庫上的具名資源方法來套用 Lake Formation 許可，並將其共用到多個 AWS Organizations、 AWS 帳戶和組織單位 (OUs)。您也可以直接與來自另一個帳戶的 IAM 主體共用聯合資料庫。

您可以使用外部 Hive 資料表上的 Lake Formation 資料篩選條件，在資料欄層級、資料列層級和儲存格層級定義精細許可。您可以使用 Amazon Athena、Amazon Redshift 或 Amazon EMR 來查詢 Lake Formation 受管外部 Hive 資料表。

如需跨帳戶資料共用和資料篩選的詳細資訊，請參閱：

- Lake Formation 中的跨帳戶資料共用

- Lake Formation 中的資料篩選和儲存格層級安全性

Data Catalog 中繼資料聯合高階步驟

1. 您可以建立具有適當許可的 IAM 使用者和角色，以部署 AWS SAM 應用程式和建立聯合資料庫。
2. 您可以透過選取使用外部 Hive 中繼存放區的資料集`Enable Data Catalog federation`選項，向 Lake Formation 註冊 Amazon S3 資料位置。
3. 您可以設定 AWS SAM 應用程式設定 (AWS Glue 連線名稱、Hive 中繼存放區的 URL 和 Lambda 函數參數），並部署 AWS SAM 應用程式。
4. AWS SAM 應用程式會部署將外部 Hive 中繼存放區與 Data Catalog 連線所需的資源。
5. 若要在 Hive 資料庫和資料表上套用 Lake Formation 許可，您可以使用 Hive 中繼存放區連線詳細資訊在 Data Catalog 中建立資料庫，並將此資料庫映射至 Hive 資料庫。
6. 將聯合資料庫的許可授予您帳戶或其他帳戶中的主體。

> **Note**
>
> 您可以將 Data Catalog 連線到外部 Hive mestastore、建立聯合資料庫，以及在 Hive 資料庫和資料表上執行查詢和 ETL 指令碼，而無需套用 Lake Formation 許可。對於未向 Lake Formation 註冊的 Amazon S3 中的來源資料，存取權取決於 Amazon S3 和 AWS Glue 動作的 IAM 許可政策。

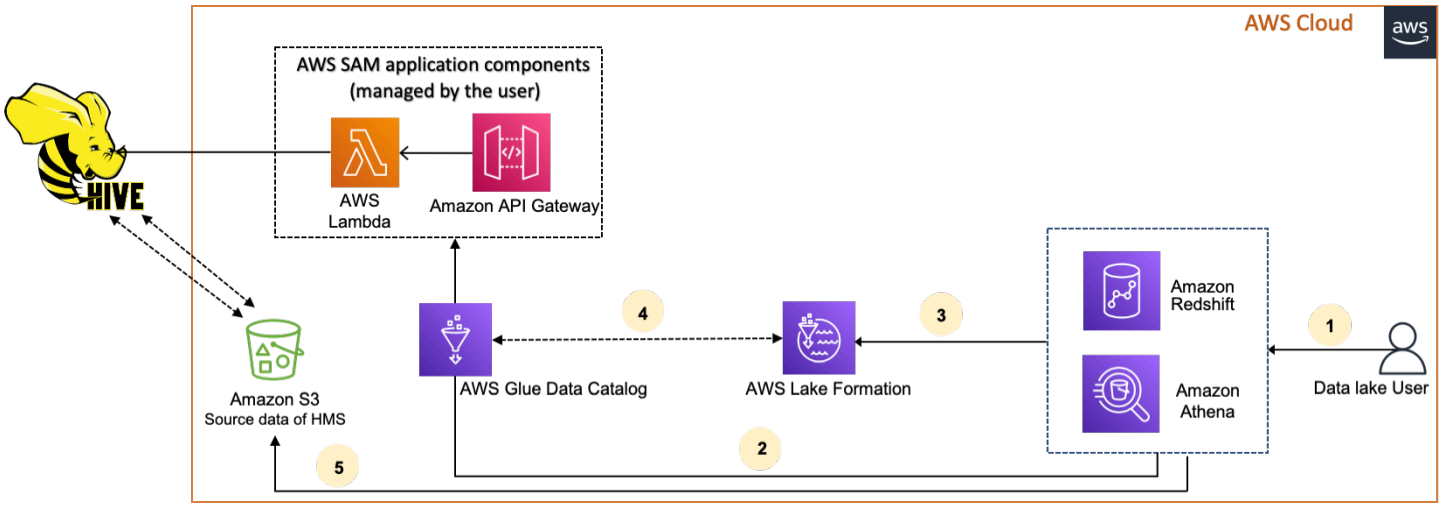
如需限制的詳細資訊，請參閱Hive 中繼資料存放區資料共用的考量和限制。

主題

- 工作流程
- 將 Data Catalog 連線至 Hive 中繼存放區的先決條件
- 將 Data Catalog 連接到外部 Hive 中繼存放區
- 其他資源

工作流程

下圖顯示將 AWS Glue Data Catalog 連接到外部 Hive 中繼存放區的工作流程。



1. 委託人使用整合的服務提交查詢，例如 Athena 或 Redshift Spectrum。
2. 整合的服務會呼叫中繼資料的資料目錄，進而呼叫 Hive 中繼存放區端點， Amazon API Gateway並接收中繼資料請求的回應。
3. 整合服務會將請求傳送至 Lake Formation，以驗證資料表資訊和登入資料來存取資料表。
4. Lake Formation 會授權請求，並將臨時登入資料提供給整合的應用程式，以允許資料存取。
5. 使用從 Lake Formation 收到的臨時登入資料，整合服務會從 Amazon S3 讀取資料，並將結果分享給委託人。

將 Data Catalog 連線至 Hive 中繼存放區的先決條件

若要將 AWS Glue Data Catalog 連接至外部 Apache Hive 中繼存放區並設定資料存取許可，您需要完成下列要求：

> **Note**
>
> 我們建議 Lake Formation 管理員部署 AWS SAM 應用程式，而且只有具有特殊權限的使用者使用 Hive 中繼存放區連線來建立對應的聯合資料庫。

1. 建立 IAM 角色。

   部署 AWS SAM 應用程式

   - 建立具有必要許可的角色，以部署建立 Hive 中繼存放區連線所需的資源 Amazon API Gateway(Lambda 函數、、IAM 角色和 AWS Glue 連線）。

建立聯合資料庫

資源需要下列許可：

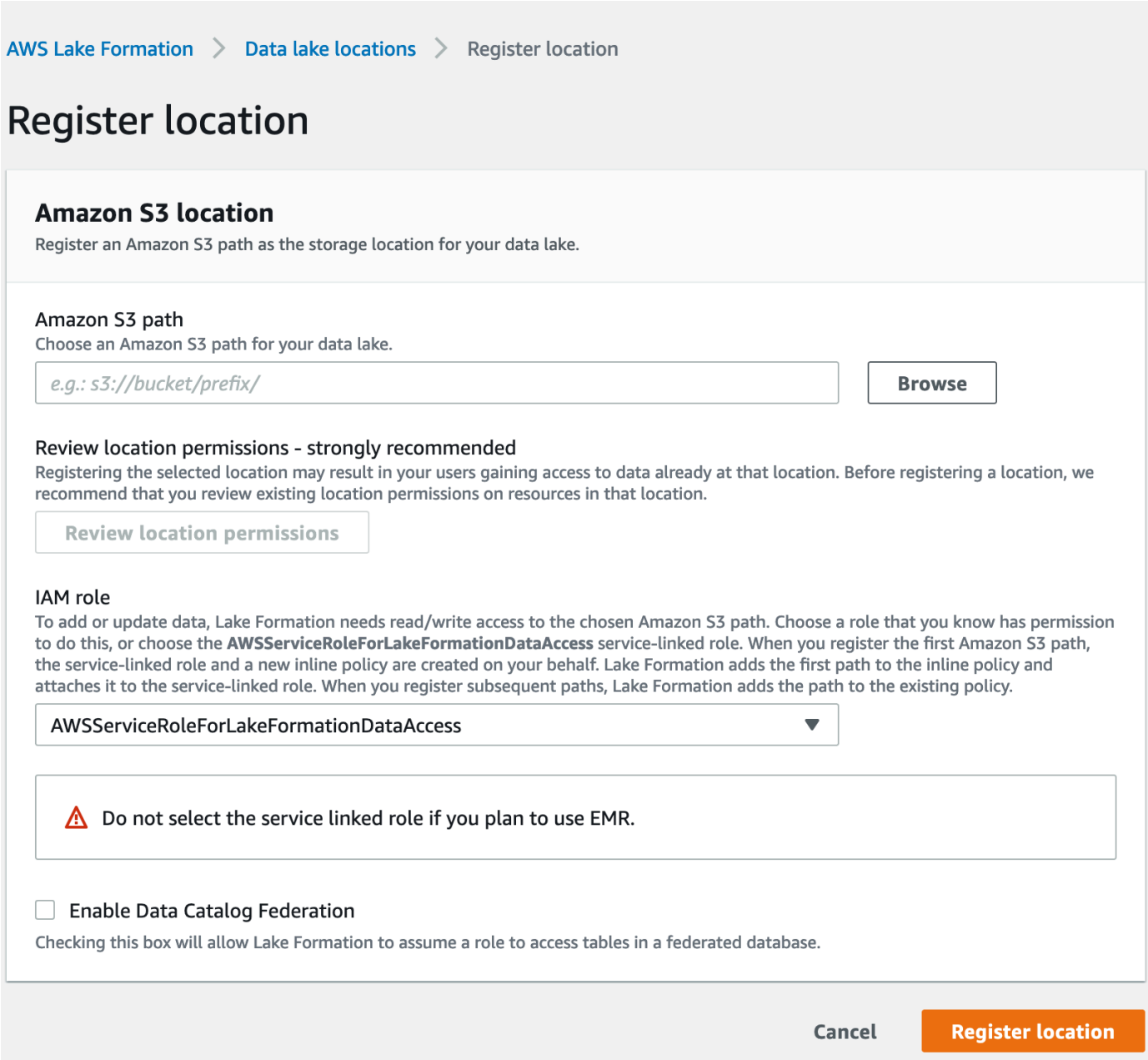
- `glue:CreateDatabase on resource arn:aws:glue:region:account-id:database/gluedatabasename`
- `glue:PassConnection on resource arn:aws:glue:region:account-id:connection/hms_connection`

2. 向 Lake Formation 註冊 Amazon S3 位置。

   若要使用 Lake Formation 來管理和保護資料湖中的資料，您必須向 Lake Formation 註冊具有 Hive 中繼存放區中資料表資料的 Amazon S3 位置。藉由這樣做，Lake Formation 可以將登入資料提供給 AWS 分析服務，例如 Athena、Redshift Spectrum 和 Amazon EMR。

   如需註冊 Amazon S3 位置的詳細資訊，請參閱 將 Amazon S3 位置新增至您的資料湖。

   當您註冊 Amazon S3 位置時，請選取啟用資料目錄聯合核取方塊，以允許 Lake Formation 擔任角色來存取聯合資料庫中的資料表。



如需向 Lake Formation 註冊資料位置的詳細資訊，請參閱 為您的資料湖設定 Amazon S3 位置。

3. 使用正確的 Amazon EMR 版本。

   若要搭配聯合 Hive 中繼存放區資料庫使用 Amazon EMR，您需要具有 Hive 3.x 版或更新版本，以及 Amazon EMR 6.x 版或更新版本。

將 Data Catalog 連接到外部 Hive 中繼存放區

若要將 AWS Glue Data Catalog 連線至 Hive 中繼存放區，您需要部署名為 GlueDataCatalogFederation-HiveMetastore AWS SAM 的應用程式。它會建立將外部 Hive 中繼存放區與 Data Catalog 連線所需的資源。您可以在 中存取 AWS SAM 應用程式 AWS Serverless Application Repository。

AWS SAM 應用程式會使用 Lambda 函數建立 Amazon API Gateway 後方 Hive 中繼存放區的連線。AWS SAM 應用程式使用統一資源識別符 (URI) 做為使用者的輸入，並將外部 Hive 中繼存放區連接到 Data Catalog。當使用者在 Hive 資料表上執行查詢時，Data Catalog 會呼叫 API Gateway 端點。端點會叫用 Lambda 函數來擷取 Hive 資料表的中繼資料。

將 Data Catalog 連線至 Hive 中繼存放區並設定許可

1. 部署 AWS SAM 應用程式。
    1. 登入 AWS Management Console 並開啟 AWS Serverless Application Repository。
    2. 選擇在導覽窗格中的 Available applications (可用的應用程式)。
    3. 選擇公有應用程式。
    4. 選取選項 Show apps that create custom IAM roles or resource policies (顯示建立自訂 IAM 角色或資源政策的應用程式)。
    5. 在搜尋方塊中，輸入名稱 GlueDataCatalogFederation-HiveMetastore。
    6. 選擇 GlueDataCatalogFederation-HiveMetastore 應用程式。
    7. 在應用程式設定下，輸入 Lambda 函數的下列最低必要設定：
        - 應用程式名稱 - 您 AWS SAM 應用程式的名稱。
        - GlueConnectionName - 連線的名稱。
        - HiveMetastoreURIs - Hive 中繼存放區主機的 URI。
        - LambdaMemory - 來自 128-10240 的 Lambda 記憶體數量，以 MB 為單位。預設值為 1024。
        - LambdaTimeout - Lambda 調用執行時間上限，以秒為單位。預設值為 30。
        - VPCSecurityGroupIds 和 VPCSubnetIds - Hive 中繼存放區所在的 VPC 資訊。
    8. 選擇 I acknowledge that this app creates custom IAM roles and resource policies (我認可此應用程式建立自訂的 IAM 角色和資源政策)。如需詳細資訊，請選擇 Info (資訊) 連結。
    9. 在 Application settings (應用程式設定) 部分的右下方，選擇 Deploy (部署)。部署完成後，Lambda 函數會出現在 Lambda 主控台的 Resources (資源) 區段中。

應用程式會部署到 Lambda。其名稱前面加上 serverlessrepo-，表示應用程式是從 部署的 AWS Serverless Application Repository。選取應用程式會帶您前往資源頁面，其中列出已部署應用程式的每個資源。資源包含 Lambda 函數，允許資料目錄與 Hive 中繼存放區、 AWS Glue 連線，以及資料庫聯合所需的其他資源之間進行通訊。

2. 在 Data Catalog 中建立聯合資料庫。

   建立 Hive 中繼存放區的連線之後，您可以在 Data Catalog 中建立指向外部 Hive 中繼存放區資料庫的聯合資料庫。您需要為連線到 Data Catalog 的每個 Hive 中繼存放區資料庫，在 Data Catalog 中建立對應的資料庫。

   Lake Formation console

   1. 在資料共用頁面上，選擇共用資料庫索引標籤，然後選擇建立資料庫。
   2. 針對連線名稱，從下拉式功能表中選擇 Hive 中繼存放區連線的名稱。
   3. 輸入資料庫的唯一資料庫名稱和聯合來源識別符。這是您在查詢資料表時在 SQL 陳述式中使用的名稱。名稱最多可包含 255 個字元，而且在您的帳戶中必須是唯一的。
   4. 選擇建立資料庫。

   AWS CLI

   ```
   aws glue create-database \
   '{
    "CatalogId": "<111122223333>",
     "database-input": {
       "Name":"<fed_glue_db>",
       "FederatedDatabase":{
           "Identifier":"<hive_db_on_emr>",
           "ConnectionName":"<hms_connection>"
        }
      }
    }'
   ```

3. 檢視聯合資料庫中的資料表。

   建立聯合資料庫之後，您可以使用 Lake Formation 主控台或 檢視 Hive 中繼存放區中的資料表清單 AWS CLI。

Lake Formation console

1. 從共用資料庫索引標籤中選取資料庫名稱。
2. 在資料庫頁面上，選擇檢視資料表。

AWS CLI

下列範例示範如何擷取連線定義、資料庫名稱，以及資料庫中的部分或全部資料表。將 Data Catalog 的 ID 取代為您用來建立資料庫的有效 AWS 帳戶 ID。將 取代hms_connection為連線名稱。

```
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. 授予許可。

   建立資料庫之後，您可以將許可授予帳戶中的其他 IAM 使用者和角色，或外部 AWS 帳戶 和組織。您將無法在聯合資料庫上授予寫入資料許可（插入、刪除）和中繼資料許可（變更、捨棄、建立）。如需授予許可的詳細資訊，請參閱 管理 Lake Formation 許可。

5. 查詢聯合資料庫。

   授予許可後，使用者可以使用 Athena 和 Amazon Redshift 登入並開始查詢聯合資料庫。使用者現在可以使用本機資料庫名稱來參考 SQL 查詢中的 Hive 資料庫。

Amazon Athena 查詢語法範例

`fed_glue_db` 將 取代為您先前建立的本機資料庫名稱。

```
Select * from fed_glue_db.customers limit 10;
```

其他資源

下列部落格文章包含設定 Hive 中繼存放區資料庫和資料表的 Lake Formation 許可，以及使用 Athena 查詢這些許可的詳細說明。我們也說明跨帳戶共用使用案例，其中生產者帳戶 A 中的 Lake Formation 主體使用 LF 標籤與消費者帳戶 B 共用聯合 Hive 資料庫和資料表。

- 使用 AWS Lake Formation 許可查詢您的 Apache Hive 中繼存放區

管理 Lake Formation 許可

Lake Formation 為資料湖中的資料提供集中存取控制。您可以在 Lake Formation 中依角色定義使用者和應用程式的安全政策型規則，並與 整合以 AWS Identity and Access Management 驗證這些使用者和角色。定義規則後，Lake Formation 會對 Amazon Redshift Spectrum 和 Amazon Athena 的使用者強制執行資料表、資料欄和 ro 層級精細度的存取控制。

主題

- 授予資料位置許可
- 授予 Data Catalog 資源的許可
- 許可範例案例
- Lake Formation 中的資料篩選和儲存格層級安全性
- 在 Lake Formation 中檢視資料庫和資料表許可
- 使用 Lake Formation 主控台撤銷許可
- Lake Formation 中的跨帳戶資料共用
- 存取和檢視共用的 Data Catalog 資料表和資料庫
- 建立資源連結
- 跨區域存取資料表

授予資料位置許可

中的資料位置許可 AWS Lake Formation 可讓主體建立和變更指向指定已註冊 Amazon S3 位置的資料目錄資源。除了 Lake Formation 資料許可之外，資料位置許可還可用於保護資料湖中的資訊。

Lake Formation 不會將 AWS Resource Access Manager (AWS RAM) 服務用於資料位置許可授予，因此您不需要接受資料位置許可的資源共用邀請。

您可以使用 Lake Formation 主控台、API 或 AWS Command Line Interface () 授予資料位置許可AWS CLI。

> **Note**
>
> 若要成功授予，您必須先向 Lake Formation 註冊資料位置。

另請參閱:

- Underlying data access control

主題

- 授予資料位置許可（相同帳戶）
- 授予資料位置許可（外部帳戶）
- 授予與您帳戶共用之資料位置的許可

授予資料位置許可（相同帳戶）

請依照下列步驟，將資料位置許可授予您 AWS 帳戶中的委託人。您可以使用 Lake Formation 主控台、 API 或 AWS Command Line Interface () 來授予許可AWS CLI。

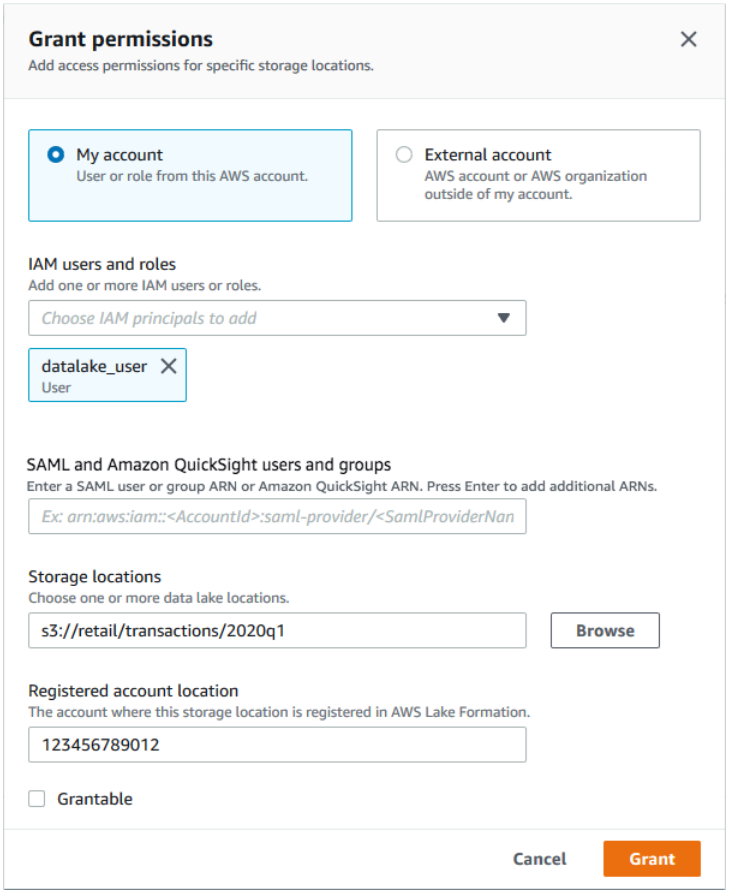
AWS Management Console

授予資料位置許可（相同帳戶）

1. 開啟 AWS Lake Formation 主控台，網址為 https：//https://console.aws.amazon.com/lakeformation/。以資料湖管理員或擁有所需資料位置之許可的委託人身分登入。
2. 在導覽窗格的許可下，選擇資料位置。
3. 選擇 Grant (授予)。
4. 在授予許可對話方塊中，確定已選取我的帳戶圖磚。然後提供下列資訊：
   - 針對 IAM 使用者和角色，選擇一或多個委託人。
   - 針對 SAML 和 Amazon QuickSight 使用者和群組，輸入一或多個 Amazon Resource Name (ARNs) 給透過 SAML 聯合的使用者或群組，或ARNs 給 Amazon QuickSight 使用者或群組。

     一次輸入一個 ARN，並在每個 ARN 之後按 Enter。如需如何建構 ARNs 的資訊，請參閱 Lake Formation 授予和撤銷 AWS CLI 命令。
   - 針對儲存位置，選擇瀏覽，然後選擇 Amazon Simple Storage Service (Amazon S3) 儲存位置。該位置必須向 Lake Formation 註冊。再次選擇瀏覽以新增其他位置。您也可以輸入位置，但請確定您在位置前面加上 `s3://`。

- 針對已註冊的帳戶位置，輸入已註冊位置 AWS 的帳戶 ID。這預設為您的帳戶 ID。在跨帳戶案例中，當將資料位置許可授予收件人帳戶中的其他主體時，收件人帳戶中的資料湖管理員可以在此處指定擁有者帳戶。
- （選用）若要讓選取的主體在選取的位置上授予資料位置許可，請選取可授予。



5. 選擇 Grant (授予)。

AWS CLI

授予資料位置許可（相同帳戶）

- 執行`grant-permissions`命令，並`DATA_LOCATION_ACCESS`授予委託人，指定 Amazon S3 路徑做為資源。

  Example

  下列範例`s3://retail`會將 上的資料位置許可授予使用者 `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

下列範例s3://retail會將 上的資料位置許可授予 ALLIAMPrincipals群組。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333:IAMPrincipals --
permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"ResourceArn":"arn:aws:s3:::retail", "CatalogId": "111122223333"}}'
```

另請參閱:

- Lake Formation 許可參考

授予資料位置許可 （外部帳戶）

請依照下列步驟，將資料位置許可授予外部 AWS 帳戶或組織。

您可以使用 Lake Formation 主控台、API 或 AWS Command Line Interface () 來授予許可AWS CLI。

開始之前

確定符合所有跨帳戶存取先決條件。如需詳細資訊，請參閱先決條件。

AWS Management Console

授予資料位置許可 （外部帳戶、主控台）

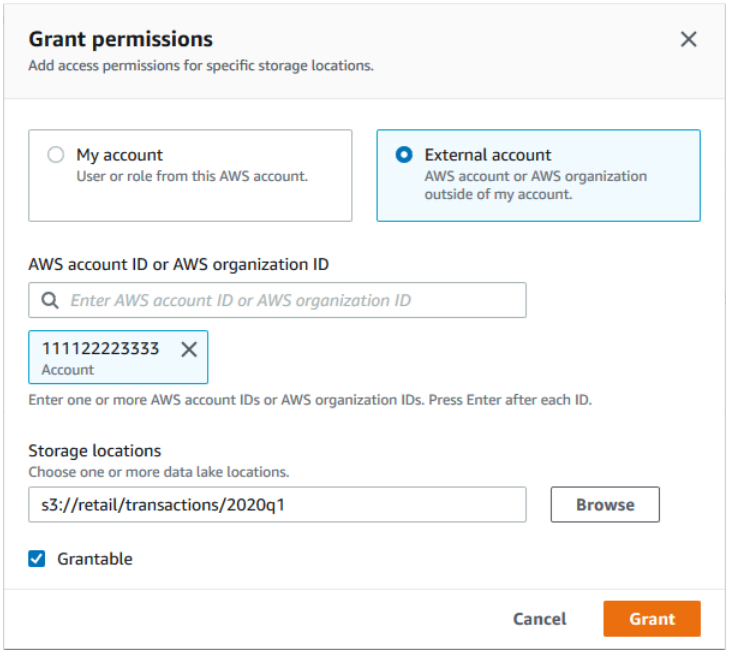
1. 開啟 AWS Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/：//。以資料湖管理員身分登入。
2. 在導覽窗格的許可下，選擇資料位置，然後選擇授予。
3. 在授予許可對話方塊中，選擇外部帳戶圖磚。
4. 請提供下列資訊：

- 針對AWS 帳戶 ID 或 AWS 組織 ID，輸入有效的 AWS 帳戶號碼、組織 IDs或組織單位 IDs。

  在每個 ID 之後按 Enter。

  組織 ID 包含「o-」，後面接著 10 到 32 個小寫字母或數字。

  組織單位 ID 包含「ou-」，後面接著 4 到 32 個小寫字母或數字（包含 OU 的根 ID)。此字串後面接著第二個 "-"（連字號）和 8 到 32 個額外的小寫字母或數字。
- 在儲存位置下，選擇瀏覽，然後選擇 Amazon Simple Storage Service (Amazon S3) 儲存位置。該位置必須向 Lake Formation 註冊。



5. 選取可授予。
6. 選擇 Grant (授予)。

AWS CLI

授予資料位置許可（外部帳戶 AWS CLI)

- 若要將許可授予外部 AWS 帳戶，請輸入類似如下的命令。

```
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"}}'
```

此命令DATA_LOCATION_ACCESS會授予 授予選項，以在 Amazon S3 位置 上帳戶 1111-2222-3333s3://retail/transactions/2020q1，該位置為帳戶 1234-5678-9012 所擁有。

若要將許可授予組織，請輸入類似如下的命令。

```
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"}}'
```

此命令DATA_LOCATION_ACCESS會將授予選項授予 Amazon S3 位置 o-abcdefghijkl上的組織s3://retail/transactions/2020q1，該位置為帳戶 1234-5678-9012 所擁有。

若要將許可授予外部 AWS 帳戶中的委託人，請輸入類似以下的命令。

```
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"}}'
```

此命令DATA_LOCATION_ACCESS會授予 Amazon S3 位置 上帳戶 1111-2222-3333 中的委託人s3://retail/transactions/2020q1，該位置為帳戶 1234-5678-9012 所擁有。

Example

下列範例s3://retail會將 上的資料位置許可授予外部帳戶中的 ALLIAMPrincipals 群組。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333:IAMPrincipals --
```

```
permissions "DATA_LOCATION_ACCESS" --resource '{ "DataLocation":
 {"ResourceArn":"arn:aws:s3:::retail", "CatalogId": "123456789012"}}'
```

另請參閱:

- Lake Formation 許可參考

授予與您帳戶共用之資料位置的許可

與 AWS 您的帳戶共用 Data Catalog 資源後，身為資料湖管理員，您可以將資源的許可授予帳戶中的其他主體。如果在共用資料表上授予ALTER許可，且資料表指向已註冊的 Amazon S3 位置，則您還必須在位置上授予資料位置許可。同樣地，如果在共用資料庫上授予 CREATE_TABLE或 ALTER許可，且資料庫具有指向已註冊位置的位置屬性，您也必須授予該位置的資料位置許可。

若要將共用位置上的資料位置許可授予您帳戶中的委託人，您的帳戶必須已使用 授予選項授予該位置的DATA_LOCATION_ACCESS許可。然後，當您DATA_LOCATION_ACCESS將 授予帳戶中的另一個委託人時，您必須包含擁有者帳戶的資料目錄 ID (AWS 帳戶 ID)。擁有者帳戶是註冊位置的帳戶。

您可以使用 AWS Lake Formation 主控台、 API 或 AWS Command Line Interface (AWS CLI 授予資料位置許可。

授予與您的帳戶共用之資料位置的許可 （主控台）

- 請遵循 授予資料位置許可 （相同帳戶） 中的步驟。

  對於儲存位置，您必須輸入位置。針對註冊帳戶位置，輸入擁有者 AWS 帳戶的帳戶 ID。

授予與您的帳戶共用之資料位置的許可 (AWS CLI)

- 輸入下列其中一個命令，將許可授予使用者或角色。

```
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>"}}'
```

授予 Data Catalog 資源的許可

您可以在 中將資料許可授予委託人， AWS Lake Formation 以便委託人可以建立和管理 Data Catalog 資源，以及存取基礎資料。您可以授予目錄、資料庫、資料表和檢視的資料湖許可。當您授予資料表的許可時，您可以限制對特定資料表資料欄或資料列的存取，以獲得更精細的存取控制。

您可以授予個別目錄、資料庫、資料表和檢視的許可，或使用單一授予操作，您可以授予目錄或資料庫中所有資料庫、資料表和檢視的許可。如果您將資料庫中所有資料表的許可授予 IAM 主體，則會隱含地授予資料庫的DESCRIBE許可。然後，資料庫會出現在 主控台的資料庫頁面上，並由 `GetDatabases` API 操作傳回。相同的原則適用於目錄層級 - 當您在目錄中接收資料庫的許可時，您也會取得該目錄的DESCRIBE許可。

> **Important**
>
> 隱含DESCRIBE許可僅適用於將許可授予相同 AWS 帳戶中的 IAM 主體時。對於跨帳戶資源，您必須明確授予DESCRIBE許可。使用屬性型存取控制 (ABAC) 時，自動DESCRIBE許可授予不適用。使用屬性授予資料庫中所有資料表的許可時，Lake Formation 不會隱含地授予資料庫DESCRIBE許可。

您可以使用具名資源方法或 Lake Formation 標籤型存取控制 (LF-TBAC) 方法來授予許可。

您可以授予許可給相同 中的委託人， AWS 帳戶 或外部帳戶或組織。當您授予外部帳戶或組織時，您會與這些帳戶或組織共用您擁有的 Data Catalog 物件。然後，這些帳戶或組織中的主體可以存取您擁有的 Data Catalog 物件和基礎資料。

> **Note**
>
> 目前，LF-TBAC 方法支援將跨帳戶許可授予 IAM 主體 AWS 帳戶、組織和組織單位 (OUs)。

當您將許可授予外部帳戶或組織時，您必須包含授予選項。只有外部帳戶中的資料湖管理員可以存取共用物件，直到管理員將共用物件的許可授予外部帳戶中的其他主體為止。

您可以使用 AWS Lake Formation 主控台、 API 或 () 授予 Data Catalog AWS Command Line Interface 許可AWS CLI。

> **Note**
>
> 當您刪除 Data Catalog 物件時，與該物件相關聯的所有許可都會失效。使用相同名稱重新建立相同的資源， 不會復原 Lake Formation 許可。使用者必須再次設定新的許可。

> **另請參閱：**
>
> - 跨 AWS 帳戶共用 Data Catalog 資料表和資料庫
> - 中繼資料存取控制
> - Lake Formation 許可參考

授予或撤銷 Lake Formation 許可所需的 IAM 許可

所有主體，包括資料湖管理員，都需要下列 AWS Identity and Access Management (IAM) 許可，才能使用 Lake Formation API 或 授予或撤銷 AWS Lake Formation Data Catalog 許可或資料位置許可 AWS CLI：

- `lakeformation:GrantPermissions`
- `lakeformation:BatchGrantPermissions`
- `lakeformation:RevokePermissions`
- `lakeformation:BatchRevokePermissions`
- `glue:GetTable`、 `glue:GetDatabase`或 `glue:GetCatalog` 適用於您使用具名資源方法授予許可的資料表、資料庫或目錄。

> **Note**
>
> 資料湖管理員具有隱含 Lake Formation 許可，可授予和撤銷 Lake Formation 許可。但他們仍然需要 Lake Formation 授予和撤銷 API 操作的 IAM 許可。
> 具有`AWSLakeFormationDataAdmin` AWS 受管政策的 IAM 角色無法新增資料湖管理員，因為此政策包含 Lake Formation API 操作 的明確拒絕`PutDataLakeSetting`。

對於非資料湖管理員且想要使用 Lake Formation 主控台授予或撤銷許可的主體，建議使用下列 IAM 政策。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:ListPermissions",
                "lakeformation:GrantPermissions",
                "lakeformation:BatchGrantPermissions",
                "lakeformation:RevokePermissions",
                "lakeformation:BatchRevokePermissions",
                "glue:GetCatalogs",
                "glue:GetDatabases",
                "glue:SearchTables",
                "glue:GetTables",
                "glue:GetCatalog",
                "glue:GetDatabase",
                "glue:GetTable",
                "iam:ListUsers",
                "iam:ListRoles",
                "sso-directory:DescribeUser",
                "sso-directory:DescribeGroup",
                "sso:DescribeInstance"
            ],
            "Resource": "*"
        }
    ]
}
```

此政策中的所有 `glue:`和 `iam:`許可都可以在 AWS 受管政策 中使用`AWSGlueConsoleFullAccess`。

若要使用 Lake Formation 標籤型存取控制 (LF-TBAC) 授予許可，委託人需要額外的 IAM 許可。如需詳細資訊，請參閱Lake Formation 標籤型存取控制最佳實務和考量事項及Lake Formation 角色和 IAM 許可參考。

跨帳戶 許可

想要使用具名資源方法授予跨帳戶 Lake Formation 許可的使用者，也必須擁有 `AWSLakeFormationCrossAccountManager` AWS 受管政策中的許可。

資料湖管理員需要這些相同的許可才能授予跨帳戶許可，加上 AWS Resource Access Manager (AWS RAM) 許可，才能將許可授予組織。如需詳細資訊，請參閱Data Lake 管理員許可。

管理使用者

具有管理許可的主體 - 例如，具有 AdministratorAccess AWS 受管政策 - 具有授予 Lake Formation 許可和建立資料湖管理員的許可。若要拒絕使用者或角色存取 Lake Formation 管理員操作，請將管理員 API 操作的Deny陳述式附加或新增至其政策。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "lakeformation:GetDataLakeSettings",
                "lakeformation:PutDataLakeSettings"
            ],
            "Effect": "Deny",
            "Resource": [
                "*"
            ]
        }
    ]
}
```

> **Important**
>
> 若要防止使用者使用擷取、轉換和載入 (ETL) 指令碼將自己新增為管理員，請確保所有非管理員使用者和角色都無法存取這些 API 操作。AWSLakeFormationDataAdmin AWS 受管政策包含 Lake Formation API 操作的明確拒絕，PutDataLakeSetting可防止使用者新增資料湖管理員。

使用具名資源方法授予資料許可

具名 Data Catalog 資源方法是一種使用集中式方法授予物件許可 AWS Glue Data Catalog 的方法，例如目錄、資料庫、資料表、資料欄和檢視。它可讓您定義以資源為基礎的政策，以控制對資料湖中特定資源的存取。

當您使用具名資源方法授予許可時，您可以指定要授予或撤銷該資源的資源類型和許可。您也可以視需要稍後撤銷許可，藉此從相關聯的資源移除許可。

您可以使用 AWS Lake Formation 主控台、APIs或 AWS Command Line Interface () 授予許可AWS CLI。

主題

- 使用具名資源方法授予目錄許可
- 使用具名資源方法授予資料庫許可
- 使用具名資源方法授予資料表許可
- 使用具名資源方法授予檢視的許可

使用具名資源方法授予目錄許可

下列步驟說明如何使用具名資源方法授予目錄許可。

Console

使用 Lake Formation 主控台上的授予許可頁面。頁面分為以下部分：

- 委託人類型：您可以將許可授予特定委託人，或使用屬性標籤。
  - 委託人 – IAM 使用者、角色、IAM Identity Center 使用者和群組、SAML 使用者和群組、 AWS 帳戶、組織或組織單位以授予許可。

    依屬性的主體 – 從 IAMroles 或 IAM 工作階段標籤新增標籤鍵值對。具有相符屬性的委託人可以存取指定的資源。
  - LF 標籤或目錄資源 – 授予許可的目錄、資料庫、資料表、檢視或資源連結。
  - 許可 – 要授予的 Lake Formation 許可。

> **Note**
>
> 若要授予資料庫資源連結的許可，請參閱 授予資源連結許可。

1. 開啟授予許可頁面。

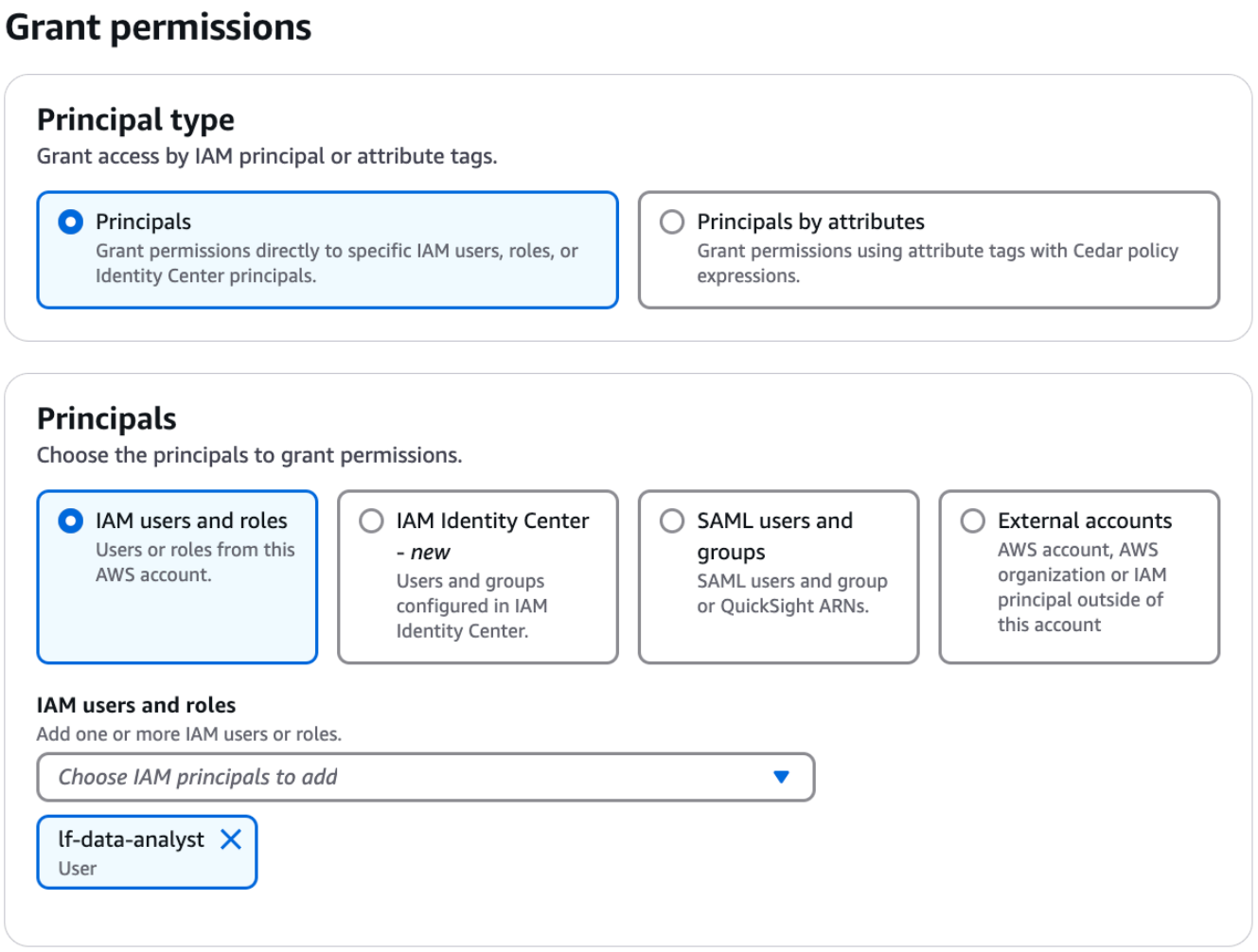
   在 https：//https://console.aws.amazon.com/lakeformation/ 開啟 AWS Lake Formation 主控台，並以資料湖管理員、目錄建立者或對目錄具有可授予許可的 IAM 使用者身分登入。

   執行以下任意一項：

- 在導覽窗格中的許可下，選擇資料許可。然後選擇授予。
- 在導覽窗格中，選擇 Data Catalog 下的目錄。 然後，在目錄頁面上，選擇目錄，然後從動作功能表的許可下，選擇授予。

> **Note**
>
> 您可以透過目錄的資源連結授予目錄許可。若要這樣做，請在目錄頁面上，選擇目錄連結容器，然後在動作功能表上，選擇對目標授予。如需詳細資訊，請參閱資源連結在 Lake Formation 中如何運作。

2. 接著，在委託人類型區段中，選擇委託人或指定連接至委託人的屬性。



指定主體

IAM 使用者和角色

從 IAM 使用者和角色清單中選擇一或多個使用者或角色。

IAM Identity Center

從使用者和群組清單中選擇一或多個使用者或群組。選取新增以新增更多使用者或群組。

SAML 使用者和群組

對於 SAML 和 QuickSight 使用者和群組，為透過 SAML 聯合的使用者或群組輸入一或多個 Amazon Resource Name (ARNs)，或為 Amazon QuickSight 使用者或群組輸入 ARNs。在每個 ARN 之後按 Enter。

如需如何建構 ARNs 的資訊，請參閱 Lake Formation 授予和撤銷 AWS CLI 命令。

> **Note**
>
> 僅 QuickSight Enterprise Edition 支援 Lake Formation 與 QuickSight 整合。

外部帳戶

針對 AWS 帳戶、 AWS organization 或 IAM Principal，輸入 IAM 使用者或角色的一或多個有效 AWS 帳戶 IDs、組織 IDs、組織單位 IDs 或 ARN。在每個 ID 之後按 Enter。

組織 ID 包含「o-」，後面接著 10-32 個小寫字母或數字。

組織單位 ID 以「ou-」開頭，後面接著 4-32 個小寫字母或數字（包含 OU 的根 ID)。此字串後面接著第二個「-」破折號和 8 到 32 個額外的小寫字母或數字。

依屬性排列的委託人

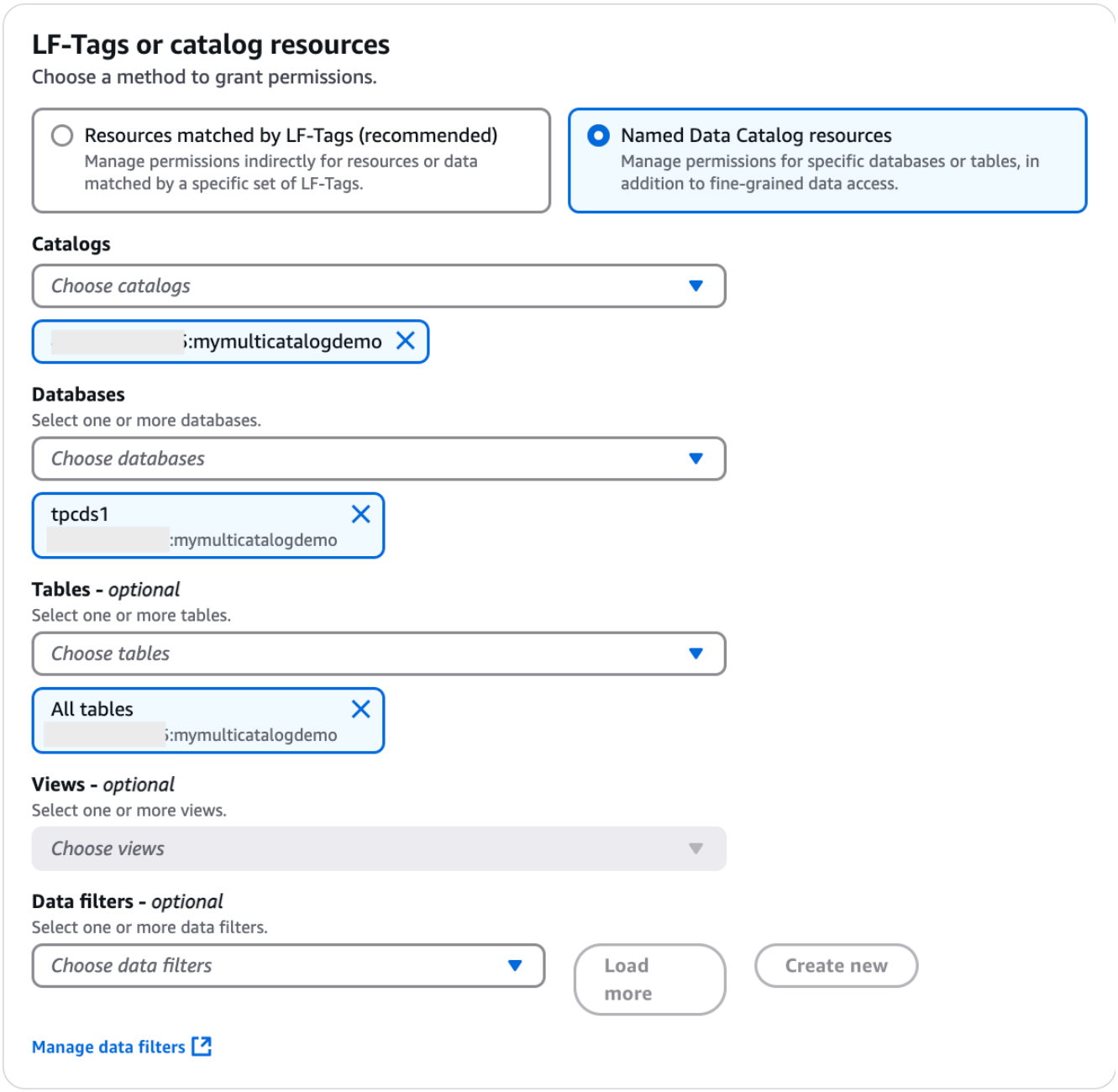
Attributes

從 IAM 角色新增 IAM 標籤鍵/值對。

許可範圍

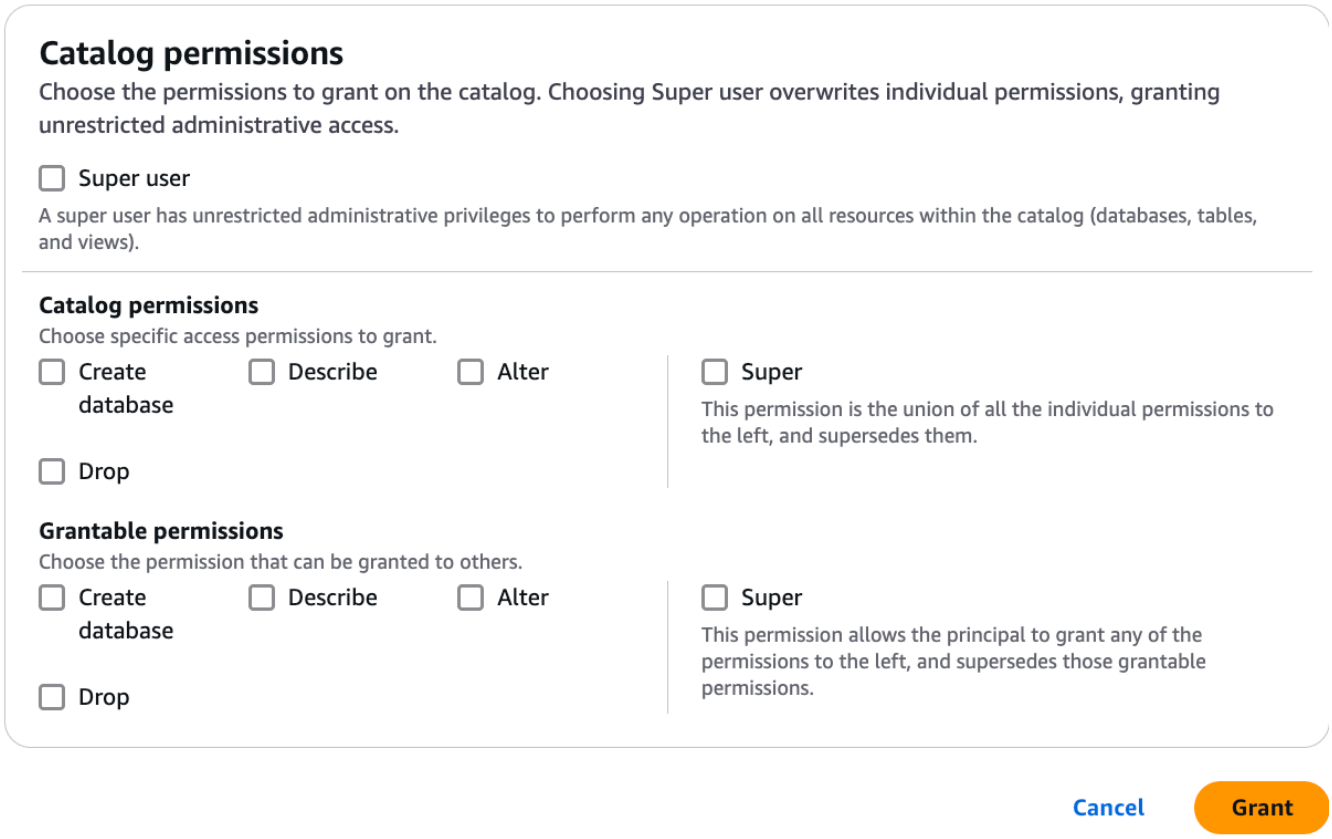
指定您是將許可授予相同帳戶或其他帳戶中具有相符屬性的委託人。

3. 在 LF 標籤或目錄資源區段中，選擇具名資料目錄資源。



4. 從目錄清單中選擇一或多個目錄。您也可以選擇一或多個資料庫、資料表和/或資料篩選條件。

5. 在目錄許可區段中，選取許可和可授予的許可。在目錄許可下，選取要授予的一或多個許可。



選擇超級使用者以授予不受限制的管理權限，對目錄中的所有資源 （資料庫、資料表和檢視） 執行任何操作。

> **Note**
>
> 在具有指向已註冊位置之位置屬性的目錄`Alter`上授予 `Create database`或 之後，請務必同時將位置上的資料位置許可授予委託人。如需詳細資訊，請參閱授予資料位置許可。

6. （選用） 在可授予許可下，選取授予收件人可以授予其 AWS 帳戶中其他委託人的許可。當您從外部帳戶將許可授予 IAM 主體時，不支援此選項。
7. 選擇 Grant (授予)。

資料許可頁面會顯示許可詳細資訊。如果您使用 Principals by 屬性選項來授予許可，您可以在清單中檢視授予 `ALLPrincipals`的許可。

AWS CLI

如需使用 授予目錄許可 AWS CLI，請參閱 建立 Amazon Redshift 聯合目錄。

使用具名資源方法授予資料庫許可

下列步驟說明如何使用具名資源方法授予資料庫許可。

Console

使用 Lake Formation 主控台上的授予許可頁面。頁面分為以下部分：

- 委託人類型 – 委託人區段包含 IAM 使用者、角色、IAM Identity Center 使用者和群組、SAML 使用者和群組、 AWS 帳戶、組織或組織單位，以授予許可。在依屬性的主體區段中，您可以指定連接到 IAM 角色之屬性的索引鍵和值。
- LF 標籤或目錄資源 – 授予許可的資料庫、資料表、檢視或資源連結。
- 許可 – 要授予的 Lake Formation 許可。

> **Note**
>
> 若要授予資料庫資源連結的許可，請參閱 授予資源連結許可。

1. 開啟授予許可頁面。

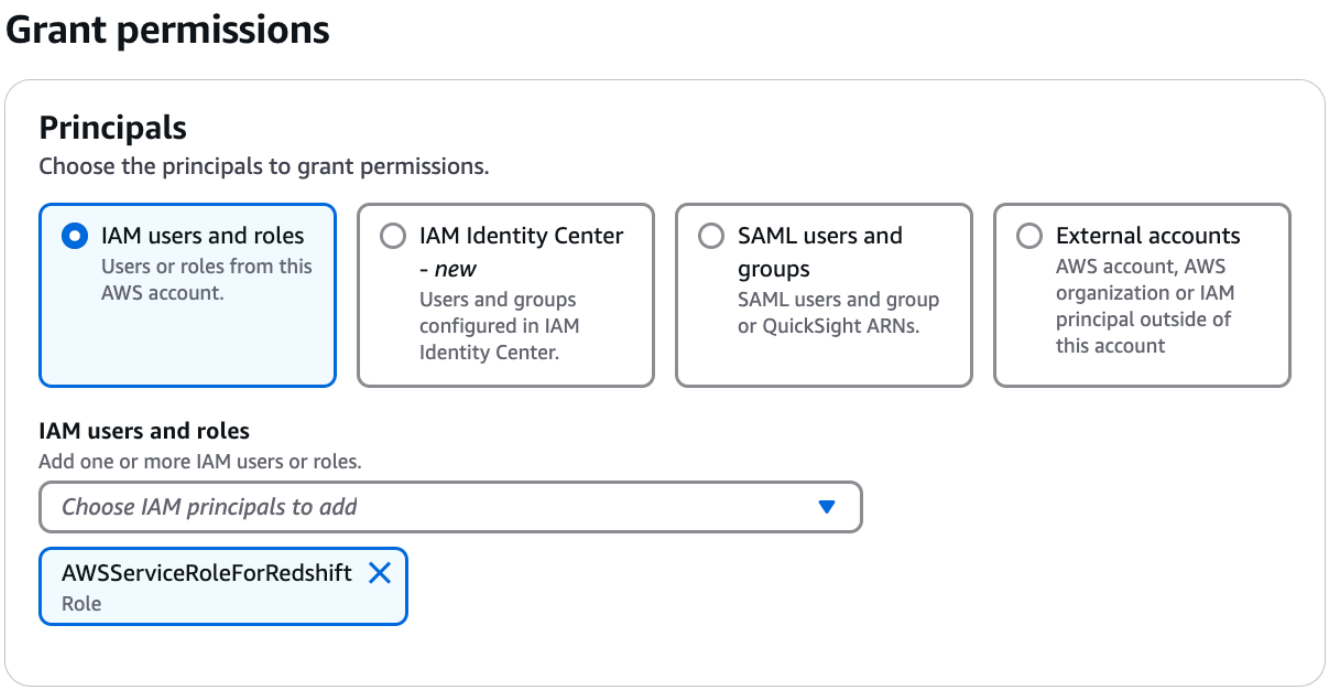
   在 https：//https://console.aws.amazon.com/lakeformation/ 開啟 AWS Lake Formation 主控台，並以資料湖管理員、資料庫建立者或對資料庫具有可授予許可的 IAM 使用者身分登入。

   執行以下任意一項：

   - 在導覽窗格中的許可下，選擇資料許可。然後選擇授予。
   - 在導覽窗格中，選擇 Data Catalog 下的資料庫。然後，在資料庫頁面上，選擇資料庫，然後從動作功能表的許可下，選擇授予。

> **Note**
>
> 您可以透過資料庫的資源連結授予資料庫的許可。若要這樣做，請在資料庫頁面上，選擇資源連結，然後在動作功能表上，選擇對目標授予。如需詳細資訊，請參閱資源連結在 Lake Formation 中如何運作。

2. 在委託人類型區段中，使用屬性指定委託人或授予許可給委託人。



IAM 使用者和角色

從 IAM 使用者和角色清單中選擇一或多個使用者或角色。

IAM Identity Center

從使用者和群組清單中選擇一或多個使用者或群組。選取新增以新增更多使用者或群組。

SAML 使用者和群組

對於 SAML 和 QuickSight 使用者和群組，為透過 SAML 聯合的使用者或群組輸入一或多個 Amazon Resource Name (ARNs)，或為 Amazon QuickSight 使用者或群組輸入 ARNs。在每個 ARN 之後按 Enter。

如需如何建構 ARNs 的資訊，請參閱 Lake Formation 授予和撤銷 AWS CLI 命令。

> **Note**
>
> 僅 QuickSight Enterprise Edition 支援 Lake Formation 與 QuickSight 整合。

外部帳戶

針對 AWS 帳戶、 AWS organization 或 IAM Principal，輸入 IAM 使用者或角色的一或多個有效 AWS 帳戶 IDs、組織 IDs、組織單位 IDs 或 ARN。在每個 ID 之後按 Enter。

組織 ID 包含「o-」，後面接著 10-32 個小寫字母或數字。

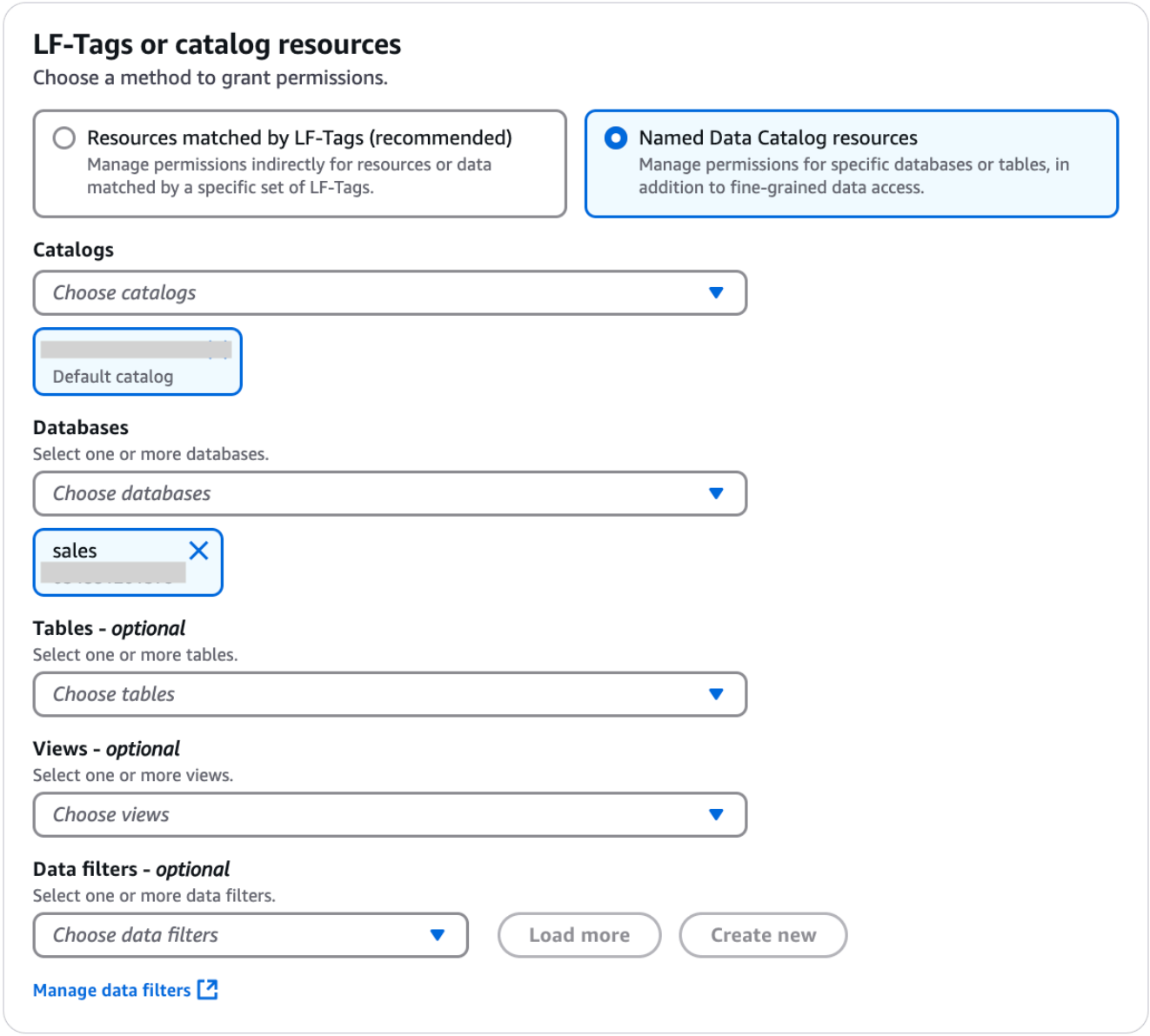
組織單位 ID 以「ou-」開頭，後面接著 4-32 個小寫字母或數字 （包含 OU 的根 ID)。此字串後面接著第二個「-」破折號和 8 到 32 個額外的小寫字母或數字。

依屬性排列的委託人

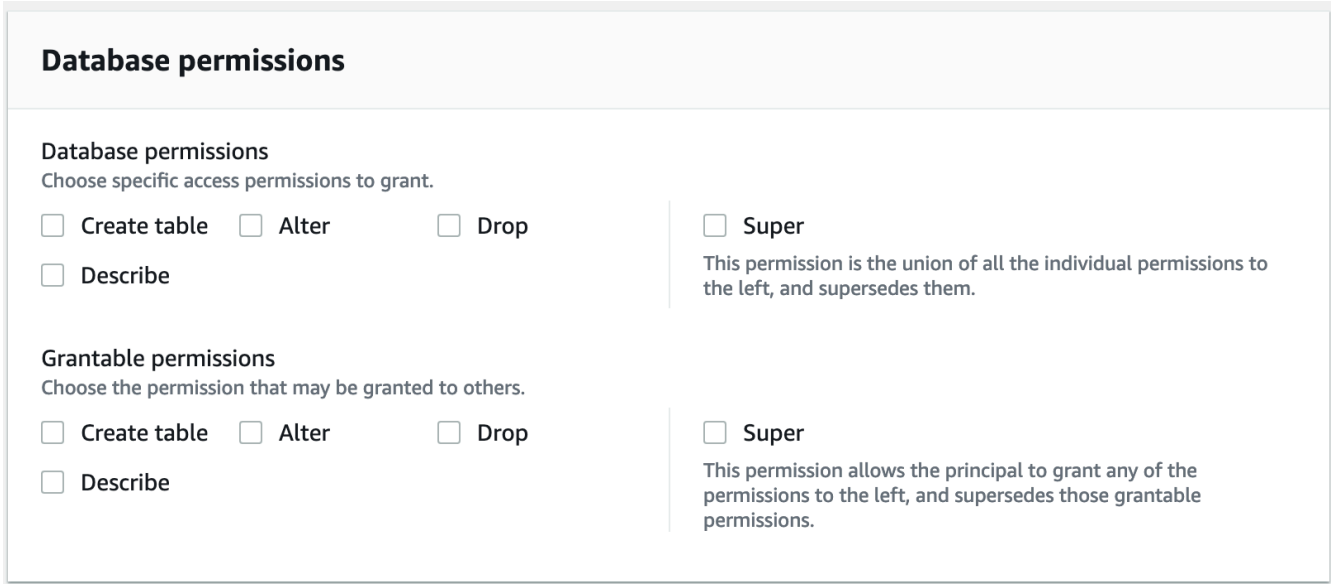
指定屬性索引鍵和值 (s)。如果您選擇多個值，則會使用 OR 運算子建立屬性表達式。這表示如果指派給 IAM 角色或使用者配對的任何屬性標籤值，角色/使用者會取得資源的存取許可。

指定您要將許可授予相同帳戶或其他帳戶中具有相符屬性的主體，以選擇許可範圍。

3. 在 LF 標籤或目錄資源區段中，選擇具名資料目錄資源。



4. 從資料庫清單中選擇一或多個資料庫。您也可以選擇一或多個資料表和/或資料篩選條件。
5. 在許可區段中，選取許可和可授予的許可。在資料庫許可下，選取要授予的一或多個許可。



> **Note**
>
> 在具有指向已註冊位置之位置屬性的資料庫`Alter`上授予 `Create Table`或 之後，請務必同時將位置上的資料位置許可授予委託人。如需詳細資訊，請參閱授予資料位置許可。

6. （選用）在可授予許可下，選取授予收件人可以授予其 AWS 帳戶中其他主體的許可。當您從外部帳戶將許可授予 IAM 主體時，不支援此選項。
7. 選擇 Grant (授予)。

AWS CLI

您可以使用具名資源方法和 AWS Command Line Interface () 授予資料庫許可AWS CLI。

使用 授予資料庫許可 AWS CLI

- 執行`grant-permissions`命令，並指定資料庫或 Data Catalog 做為資源，視授予的許可而定。

  在下列範例中，將 *<account-id>* 取代為有效的 AWS 帳戶 ID。

  Example – 授予 以建立資料庫

  此範例`CREATE_DATABASE`會授予使用者 `datalake_user1`。由於授與此許可的資源是 Data Catalog，因此 命令會指定空`CatalogResource`結構做為 `resource` 參數。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/datalake_user1 --
permissions "CREATE_DATABASE" --resource '{ "Catalog": {}}'
```

Example – 授予 在指定的資料庫中建立資料表

下一個範例會將資料庫CREATE_TABLE上的 授予retail使用者 datalake_user1。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::<account-id>:user/datalake_user1 --
permissions "CREATE_TABLE" --resource '{ "Database": {"Name":"retail"}}'
```

Example – 使用授予選項授予外部 AWS 帳戶

下一個範例CREATE_TABLE會將 資料庫上的授予選項授予retail外部帳戶 1111-2222-3333。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333 --permissions "CREATE_TABLE"
 --permissions-with-grant-option "CREATE_TABLE" --resource '{ "Database":
 {"Name":"retail"}}'
```

Example – 授予組織

下一個範例ALTER會將 資料庫上的授予選項issues授予組織 o-abcdefghijkl。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:organizations::111122223333:organization/
o-abcdefghijkl --permissions "ALTER" --permissions-with-grant-option "ALTER" --
resource '{ "Database": {"Name":"issues"}}'
```

Example - 授予相同帳戶中**ALLIAMPrincipals**的

下一個範例會將資料庫的CREATE_TABLE許可授予相同帳戶中retail的所有主體。此選項可讓帳戶中的每個主體在資料庫中建立資料表，並建立資料表資源連結，允許整合式查詢引擎存取共用資料庫和資料表。當委託人收到跨帳戶授予，而且沒有建立資源連結的許可時，此選項特別有用。在此案例中，資料湖管理員可以建立預留位置資料庫，並將CREATE_TABLE許可授予 ALLIAMPrincipal群組，讓帳戶中的每個 IAM 主體都能在預留位置資料庫中建立資源連結。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333:IAMPrincipals
 --permissions "CREATE_TABLE"  --resource '{ "Database":
 {"Name":"temp","CatalogId":"111122223333"}}'
```

Example - 授予外部帳戶中**ALLIAMPrincipals**的

下一個範例會將資料庫CREATE_TABLE上的 授予外部帳戶retail中的所有主體。此選項可讓帳戶中的每個主體在資料庫中建立資料表。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=111122223333:IAMPrincipals
 --permissions "CREATE_TABLE"  --resource '{ "Database":
 {"Name":"retail","CatalogId":"123456789012"}}'
```

> **Note**
>
> 在具有指向已註冊位置之位置屬性的資料庫ALTER上授予 CREATE_TABLE或 之後，請務必同時將位置上的資料位置許可授予委託人。如需詳細資訊，請參閱授予資料位置許可。

> **另請參閱**
>
> - Lake Formation 許可參考
> - 授予與您帳戶共用之資料庫或資料表的許可
> - 存取和檢視共用的 Data Catalog 資料表和資料庫

使用具名資源方法授予資料表許可

您可以使用 Lake Formation 主控台或 AWS CLI 授予資料目錄資料表的 Lake Formation 許可。您可以授予個別資料表的許可，或使用單一授予操作，您可以授予資料庫中所有資料表的許可。

如果您授予資料庫中所有資料表的許可，則會隱含授予資料庫的DESCRIBE許可。然後，資料庫會出現在 主控台的資料庫頁面上，並由 GetDatabases API 操作傳回。使用屬性型存取控制 (ABAC) 時，此

自動DESCRIBE許可授予不適用。使用屬性授予資料庫中所有資料表的許可時，Lake Formation 不會隱含地授予資料庫DESCRIBE許可。

當您選擇 SELECT做為授予許可時，您可以選擇套用資料欄篩選條件、資料列篩選條件或儲存格篩選條件。

Console

下列步驟說明如何使用 Lake Formation 主控台上的具名資源方法和授予資料湖許可頁面來授予資料表許可。頁面分為以下部分：

- 委託人類型 – 要授予許可的使用者、角色 AWS 、帳戶、組織或組織單位。您也可以將許可授予具有相符屬性的委託人。
- LF 標籤或目錄資源 – 要授予許可的資料庫、資料表或資源連結。
- 許可 – 要授予的 Lake Formation 許可。

> **Note**
>
> 若要授予資料表資源連結的許可，請參閱 授予資源連結許可。

1. 開啟授予許可頁面。

   在 https：//https://console.aws.amazon.com/lakeformation/ 開啟 AWS Lake Formation 主控台，並以資料湖管理員、資料表建立者或使用者身分登入，該使用者已獲授予 授予選項的資料表許可。

   執行以下任意一項：

   - 在導覽窗格中，選擇許可下的資料許可。然後選擇授予。
   - 在導覽窗格中，選擇 Tables (資料表)。然後，在資料表頁面上，選擇資料表，然後在動作功能表的許可下，選擇授予。

> **Note**
>
> 您可以透過資料表的資源連結授予許可。若要這樣做，請在資料表頁面上，選擇資源連結，然後在動作功能表上，選擇對目標授予。如需詳細資訊，請參閱資源連結在 Lake Formation 中如何運作。

2. 接下來，在委託人類型區段中，指定具有相符屬性的委託人或委託人，以授予許可。

IAM 使用者和角色

從 IAM 使用者和角色清單中選擇一或多個使用者或角色。

IAM Identity Center

從使用者和群組清單中選擇一或多個使用者或群組。

SAML 使用者和群組

對於 SAML 和 QuickSight 使用者和群組，輸入透過 SAML 聯合的使用者或群組的一或多個 Amazon Resource Name (ARNs)，或 QuickSight 使用者或群組ARNs。在每個 ARN 之後按 Enter。

如需如何建構 ARNs 的資訊，請參閱 Lake Formation 授予和撤銷 AWS CLI 命令。

> **Note**
>
> 僅 QuickSight Enterprise Edition 支援 Lake Formation 與 QuickSight 整合。

外部帳戶

對於 AWS 帳戶 、 AWS organization 或 IAM Principal，輸入一或多個 AWS 帳戶 IDs、組織 IDs、組織單位 IDs 或 IAM 使用者或角色的 ARN。在每個 ID 之後按 Enter。

組織 ID 包含「o-」，後面接著 10-32 個小寫字母或數字。

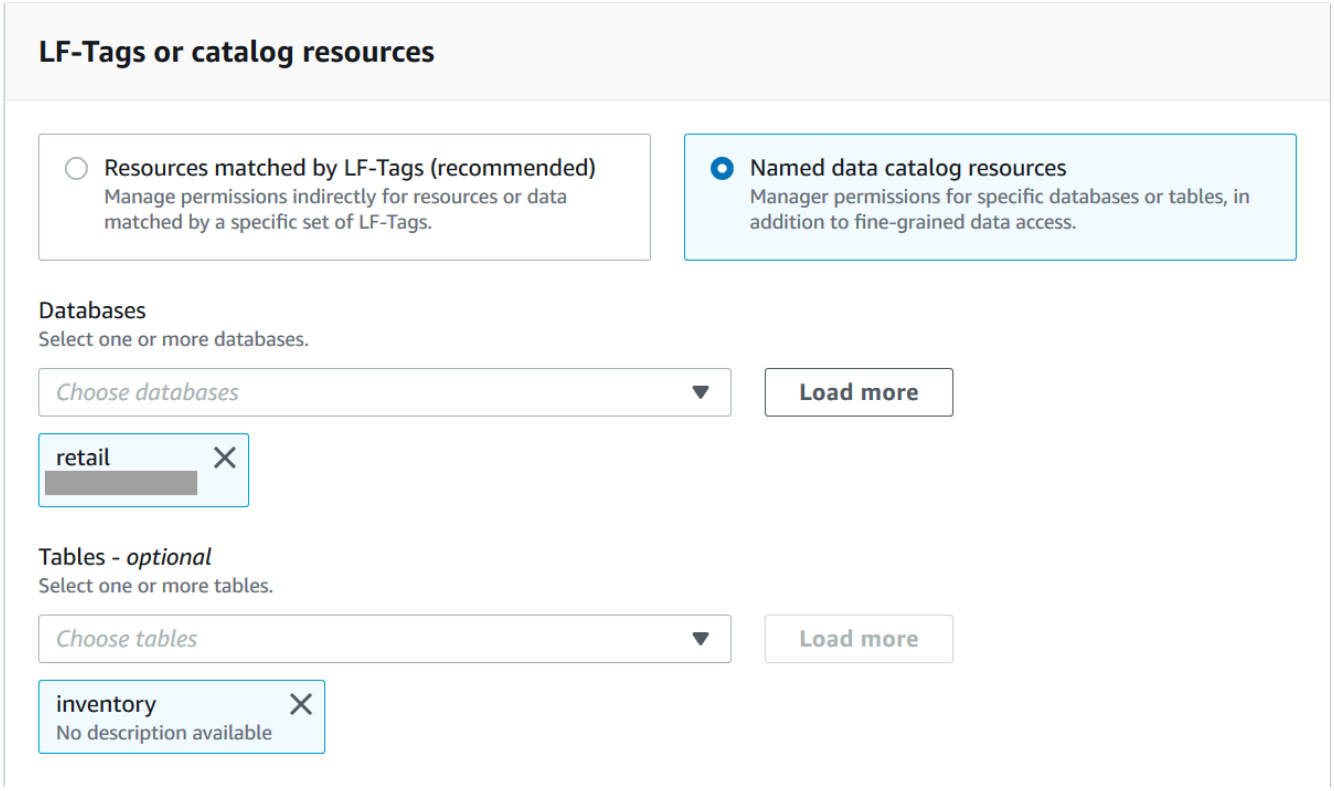
組織單位 ID 以「ou-」開頭，後面接著 4-32 個小寫字母或數字 （包含 OU 的根 ID)。此字串後面接著第二個「-」字元和 8 到 32 個額外的小寫字母或數字。

依屬性排列的委託人

指定屬性索引鍵和值 (s)。如果您選擇多個值，則會使用 OR 運算子建立屬性表達式。這表示如果指派給 IAM 角色或使用者配對的任何屬性標籤值，角色/使用者會取得資源的存取許可

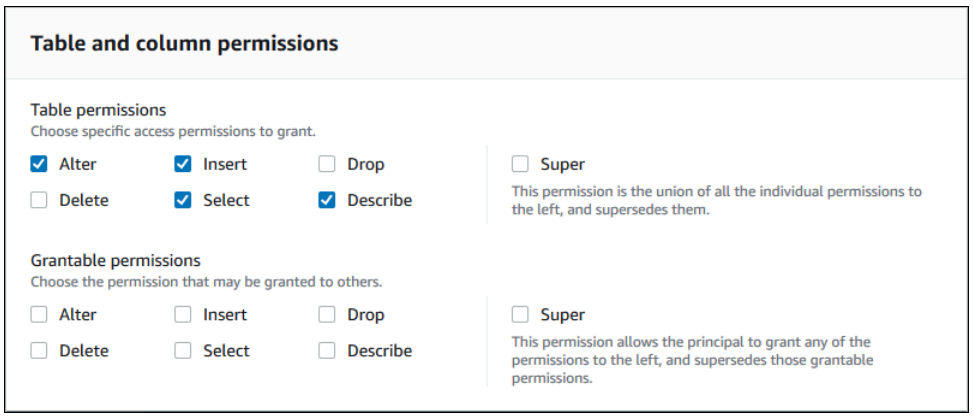
指定您要將許可授予相同帳戶或其他帳戶中具有相符屬性的主體，以選擇許可範圍。

3. 在 LF 標籤或目錄資源區段中，選擇資料庫。然後選擇一或多個資料表，或所有資料表。

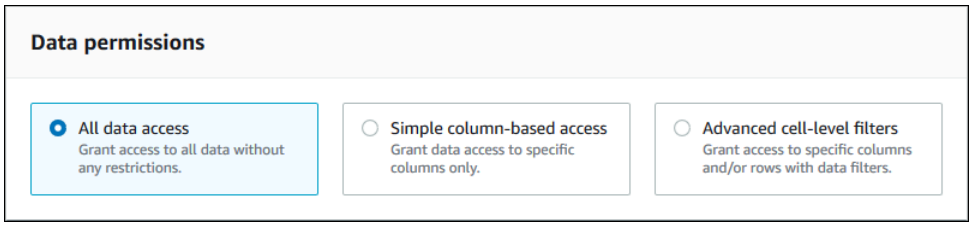


4. 指定沒有資料篩選的許可。

在許可區段中，選取要授予的資料表許可，並選擇性選取可授予的許可。



如果您授予選取，資料許可區段會出現在資料表和資料欄許可區段下方，並預設為選取所有資料存取選項。接受預設值。



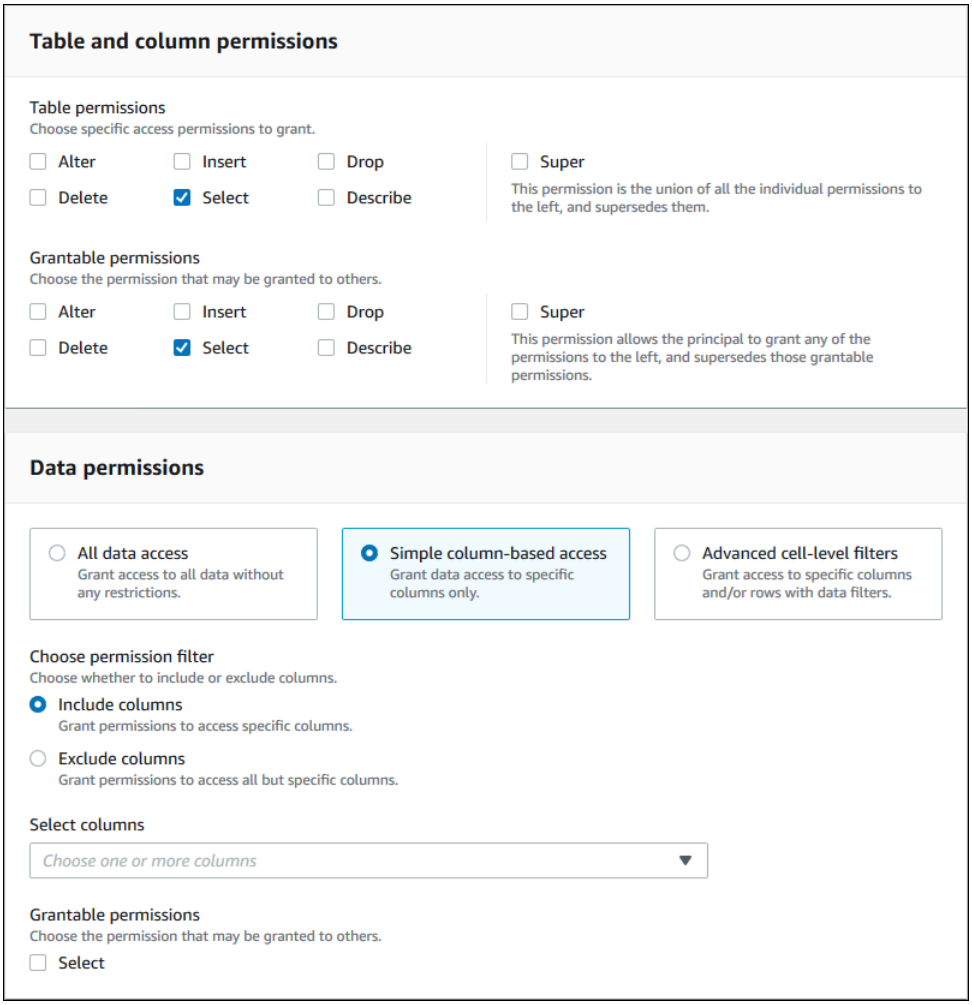
5. 選擇 Grant (授予)。
6. 使用資料篩選指定選取許可

   選取選取許可。請勿選取任何其他許可。

   資料許可區段會出現在資料表和資料欄許可區段下方。
7. 執行以下任意一項：

   - 僅套用簡單的資料欄篩選。

     1. 選擇簡易資料欄型存取。



2. 選擇是否包含或排除資料欄，然後選擇要包含或排除的資料欄。

   將許可授予外部 AWS 帳戶或組織時，僅支援包含清單。

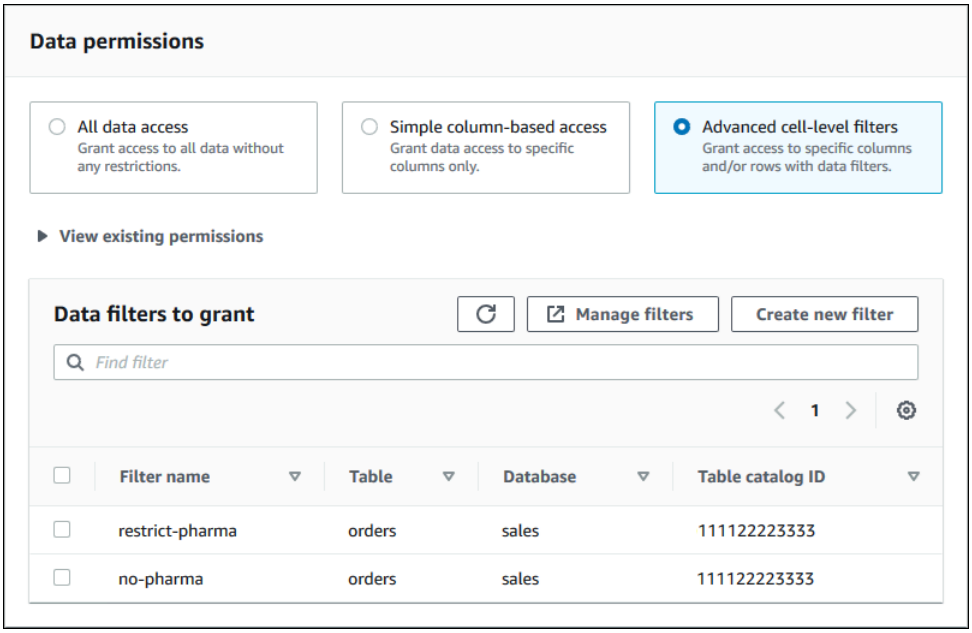
3. （選用）在可授予許可下，開啟選取許可的授予選項。

   如果您包含授予選項，則授予收件人只能在您授予給他們的資料欄上授予許可。

> **Note**
>
> 您也可以透過建立資料篩選條件來套用資料欄篩選，以指定資料欄篩選條件，並將所有資料列指定為資料列篩選條件。不過，這需要更多步驟。

- 套用資料欄、資料列或儲存格篩選。

  1. 選擇進階儲存格層級篩選條件。



2. （選用）展開 檢視現有的許可。
3. （選用）選擇建立新篩選條件。
4. （選用）若要檢視列出的篩選條件的詳細資訊，或建立新的或刪除現有的篩選條件，請選擇管理篩選條件。

   資料篩選條件頁面會在新的瀏覽器視窗中開啟。

   當您在資料篩選條件頁面上完成時，請返回授予許可頁面，並視需要重新整理頁面以檢視您建立的任何新資料篩選條件。
5. 選取要套用至授予的一或多個資料篩選條件。

> **Note**
>
> 如果清單中沒有資料篩選條件，表示未為選取的資料表建立資料篩選條件。

8. 選擇 Grant (授予)。

AWS CLI

您可以使用具名資源方法和 () AWS Command Line Interface 授予資料表許可AWS CLI。

使用 授予資料表許可 AWS CLI

- 執行`grant-permissions`命令，並將資料表指定為資源。

Example – 授予單一資料表 - 不篩選

下列範例ALTER會將 SELECT和 AWS 授予資料庫 datalake_user1 中 資料表的帳戶 1111-2222-3333 inventory中的使用者retail。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1 --
permissions "SELECT" "ALTER" --resource '{ "Table": {"DatabaseName":"retail",
 "Name":"inventory"}}'
```

> **Note**
>
> 如果您在已註冊位置對其基礎資料的資料表上授予 ALTER許可，請務必同時將位置上的資料位置許可授予委託人。如需詳細資訊，請參閱授予資料位置許可。

Example – 使用授予選項授予所有資料表 - 不篩選

下一個範例SELECT授予資料庫 中所有資料表的授予選項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 – 使用簡單的資料欄篩選授予

下一個範例SELECT會授予資料表 中的資料欄子集persons。它使用簡單的資料欄篩選。

```
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 – 使用資料篩選條件授予

此範例會授予 orders資料表SELECT，並套用restrict-pharma資料篩選條件。

```
aws lakeformation grant-permissions --cli-input-json file://grant-params.json
```

以下是 檔案 的內容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"]
}
```

另請參閱

- Lake Formation 許可概觀
- Lake Formation 中的資料篩選和儲存格層級安全性
- Lake Formation 角色和 IAM 許可參考
- 授予資源連結許可
- 存取和檢視共用的 Data Catalog 資料表和資料庫

使用具名資源方法授予檢視的許可

下列步驟說明如何使用具名資源方法和授予許可頁面來授予檢視的許可。頁面分為以下部分：

- 委託人類型 – 授予許可的 IAM 使用者、角色、IAM Identity Center 使用者和群組 AWS 帳戶、組織或組織單位。您也可以將許可授予具有相符屬性的主體。
- LF 標籤或目錄資源 – 授予許可的資料庫、資料表、檢視或資源連結。
- 許可 – 要授予的資料湖許可。

開啟授予許可頁面

1. 在 https：//https://console.aws.amazon.com/lakeformation/ 開啟 AWS Lake Formation 主控台，並以資料湖管理員、資料庫建立者或對資料庫具有可授予許可的 IAM 使用者身分登入。
2. 執行以下任意一項：

   - 在導覽窗格中的許可下，選擇資料許可。然後選擇授予。
   - 在導覽窗格中，選擇 Data Catalog 下的檢視。然後，在檢視頁面上，選擇檢視，然後從動作功能表的許可下，選擇授予。

> **Note**
>
> 您可以透過檢視的資源連結授予檢視許可。若要這樣做，請在檢視頁面上，選擇資源連結，然後在動作功能表上，選擇對目標授予。如需詳細資訊，請參閱資源連結在 Lake Formation 中如何運作。

指定委託人類型

在主體類型區段中，依屬性選擇主體或主體。如果您選擇主體，可使用下列選項：

IAM 使用者和角色

從 IAM 使用者和角色清單中選擇一或多個使用者或角色。

IAM Identity Center

從使用者和群組清單中選擇一或多個使用者或群組。

SAML 使用者和群組

對於 SAML 和 QuickSight 使用者和群組，為透過 SAML 聯合的使用者或群組輸入一或多個 Amazon Resource Name (ARNs)，或為 Amazon QuickSight 使用者或群組輸入 ARNs。在每個 ARN 之後按 Enter。

如需如何建構 ARNs 的資訊，請參閱 Lake Formation 授予和撤銷 AWS CLI 命令。

> **Note**
>
> 僅 QuickSight Enterprise Edition 支援 Lake Formation 與 QuickSight 整合。

外部帳戶

針對 AWS 帳戶、 AWS organization 或 IAM Principal，輸入 IAM 使用者或角色的一或多個有效 AWS 帳戶 IDs、組織 IDs、組織單位 IDs 或 ARN。在每個 ID 之後按 Enter。

組織 ID 包含「o-」，後面接著 10-32 個小寫字母或數字。

組織單位 ID 以「ou-」開頭，後面接著 4-32 個小寫字母或數字 （包含 OU 的根 ID)。此字串後面接著第二個「-」破折號和 8 到 32 個額外的小寫字母或數字。

> 另請參閱
>
> - 存取和檢視共用的 Data Catalog 資料表和資料庫

依屬性排列的委託人

指定屬性索引鍵和值 (s)。如果您選擇多個值，則會使用 OR 運算子建立屬性表達式。這表示如果指派給 IAM 角色或使用者配對的任何屬性標籤值，該角色/使用者會取得資源的存取許可

指定您要將許可授予相同帳戶或其他帳戶中具有相符屬性的主體，以選擇許可範圍。

指定檢視

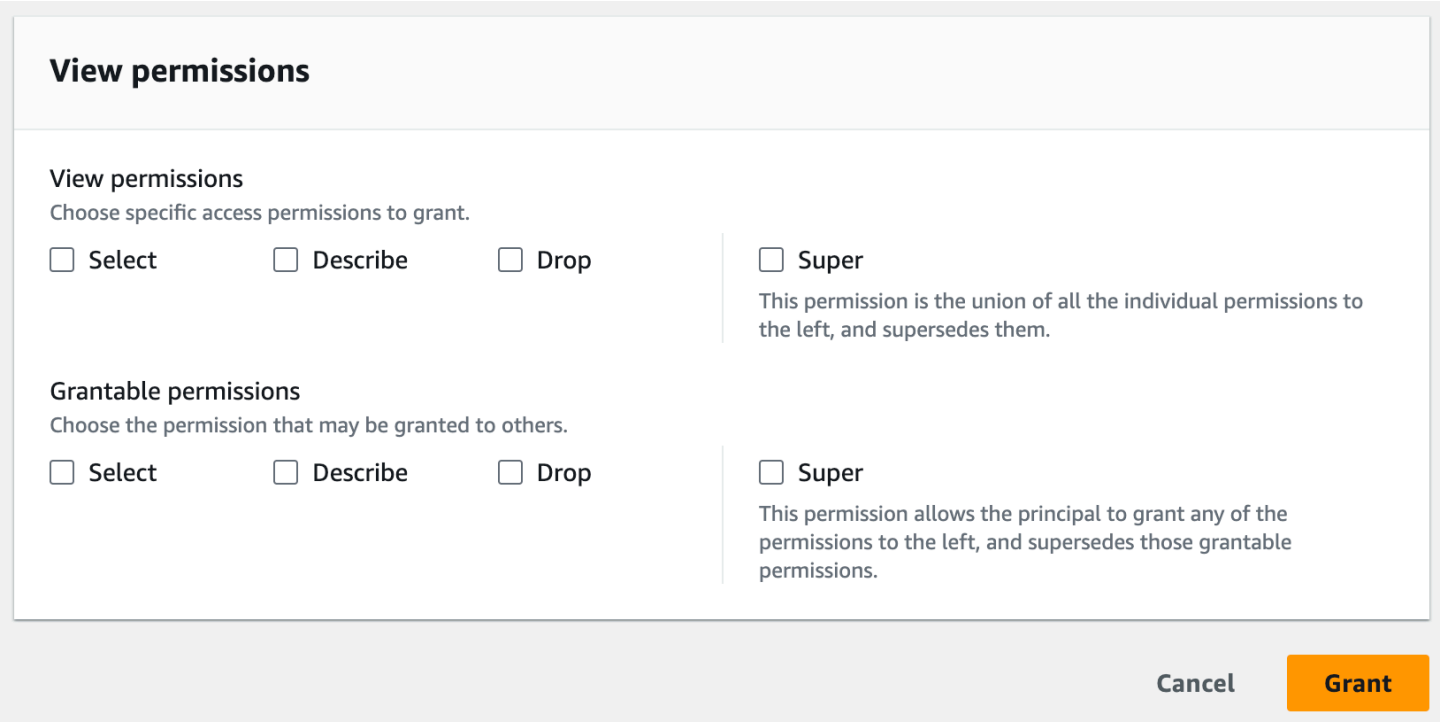
在 LF 標籤或目錄資源區段中，選擇要授予許可的一或多個檢視。

1. 選擇具名資料目錄資源。
2. 從檢視清單中選擇一或多個檢視。您也可以選擇一或多個目錄、資料庫、資料表和/或資料篩選條件。

   將資料湖許可授予資料庫中`All tables`的 ，會導致承授者擁有資料庫中所有資料表和檢視的許可。

指定許可

在許可區段中，選取許可和可授予的許可。



1. 在檢視許可下，選取要授予的一或多個許可。
2. （選用）在可授予許可下，選取授予收件人可以在其 中授予其他委託人的許可 AWS 帳戶。當您從外部帳戶將許可授予 IAM 主體時，不支援此選項。
3. 選擇 Grant (授予)。

另請參閱

- Lake Formation 許可參考
- 授予與您帳戶共用之資料庫或資料表的許可

Lake Formation 標籤型存取控制

Lake Formation 標籤型存取控制 (LF-TBAC) 是一種授權策略，可根據屬性定義許可。在 Lake Formation 中，這些屬性稱為 LF 標籤。您可以將 LF 標籤連接至 Data Catalog 資源，並使用這些 LF 標籤將許可授予 Lake Formation 主體。當委託人的標籤值符合資源標籤值時，Lake Formation 允許對這些資源進行操作。LF-TBAC 有助於快速成長的環境，並有助於處理政策管理變得繁瑣的情況。

當有大量 Data Catalog 資源時，建議使用 LF-TBAC 來授予 Lake Formation 許可。LF-TBAC 比具名資源方法更具可擴展性，且需要較少的許可管理開銷。

> **Note**
>
> IAM 標籤與 LF 標籤不同。這些標籤不可互換。LF 標籤用於授予 Lake Formation 許可，而 IAM 標籤用於定義 IAM 政策。

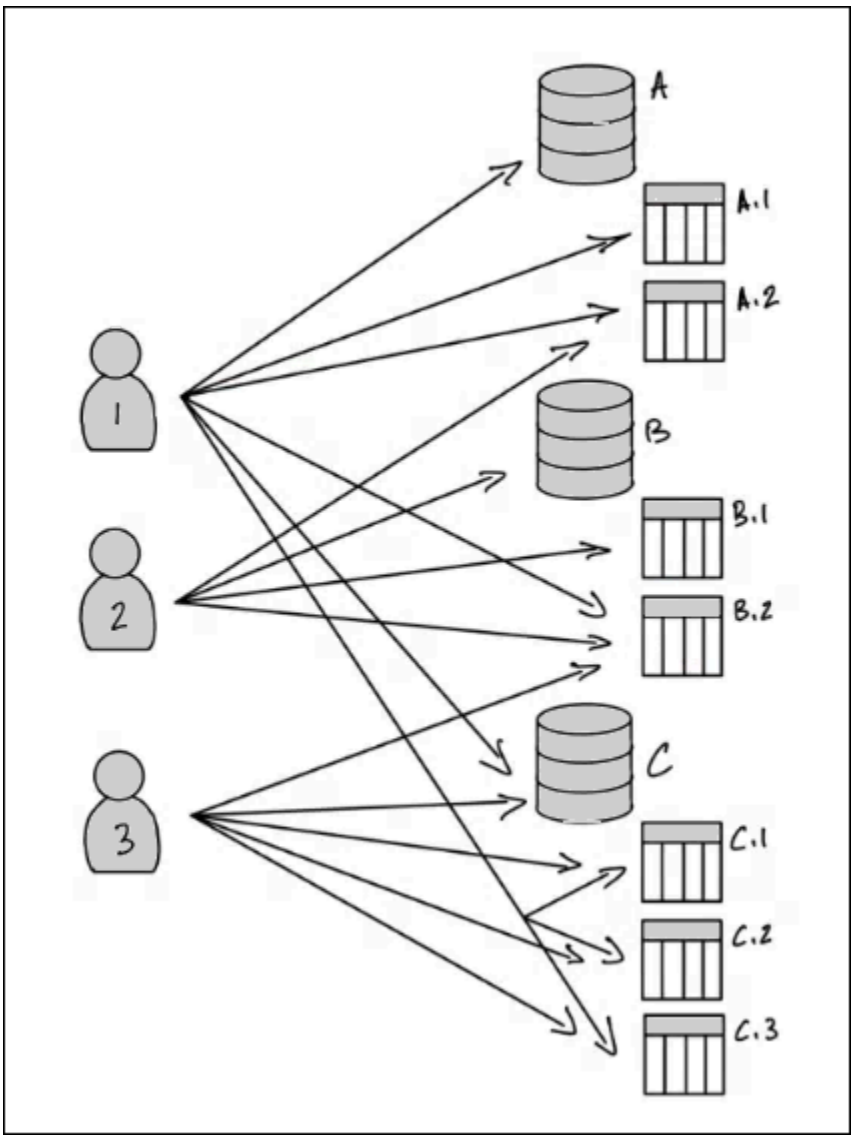
Lake Formation 標籤型存取控制的運作方式

每個 LF-Tag 都是金鑰值對，例如 `department=sales`或 `classification=restricted`。金鑰可以有多個定義的值，例如 `department=sales,marketing,engineering,finance`。

若要使用 LF-TBAC 方法，資料湖管理員和資料工程師會執行下列任務。

任務	任務詳細資訊
1. 定義 LF 標籤的屬性和關係。	-
2. 在 Lake Formation 中建立 LF 標籤建立者。	新增 LF 標籤建立者
3. 在 Lake Formation 中建立 LF 標籤。	建立 LF 標籤
4. 將 LF 標籤指派給 Data Catalog 資源。	將 LF 標籤指派給 Data Catalog 資源
5. 將許可授予其他主體，以將 LF 標籤指派給資源，選擇性使用授予選項。	管理 LF-Tag 值許可
6. 將 LF-Tag 表達式授予委託人，選擇性使用授予選項。	使用 LF-TBAC 方法授予資料湖許可
7.（建議）驗證委託人是否可透過 LF-TBAC 方法存取正確的資源後，請撤銷使用具名資源方法授予的許可。	-

假設您必須將許可授予三個資料庫和七個資料表上的三個主體。



若要使用具名資源方法達成上圖中指定的許可，您必須進行 17 次授予，如下所示（虛擬程式碼）。

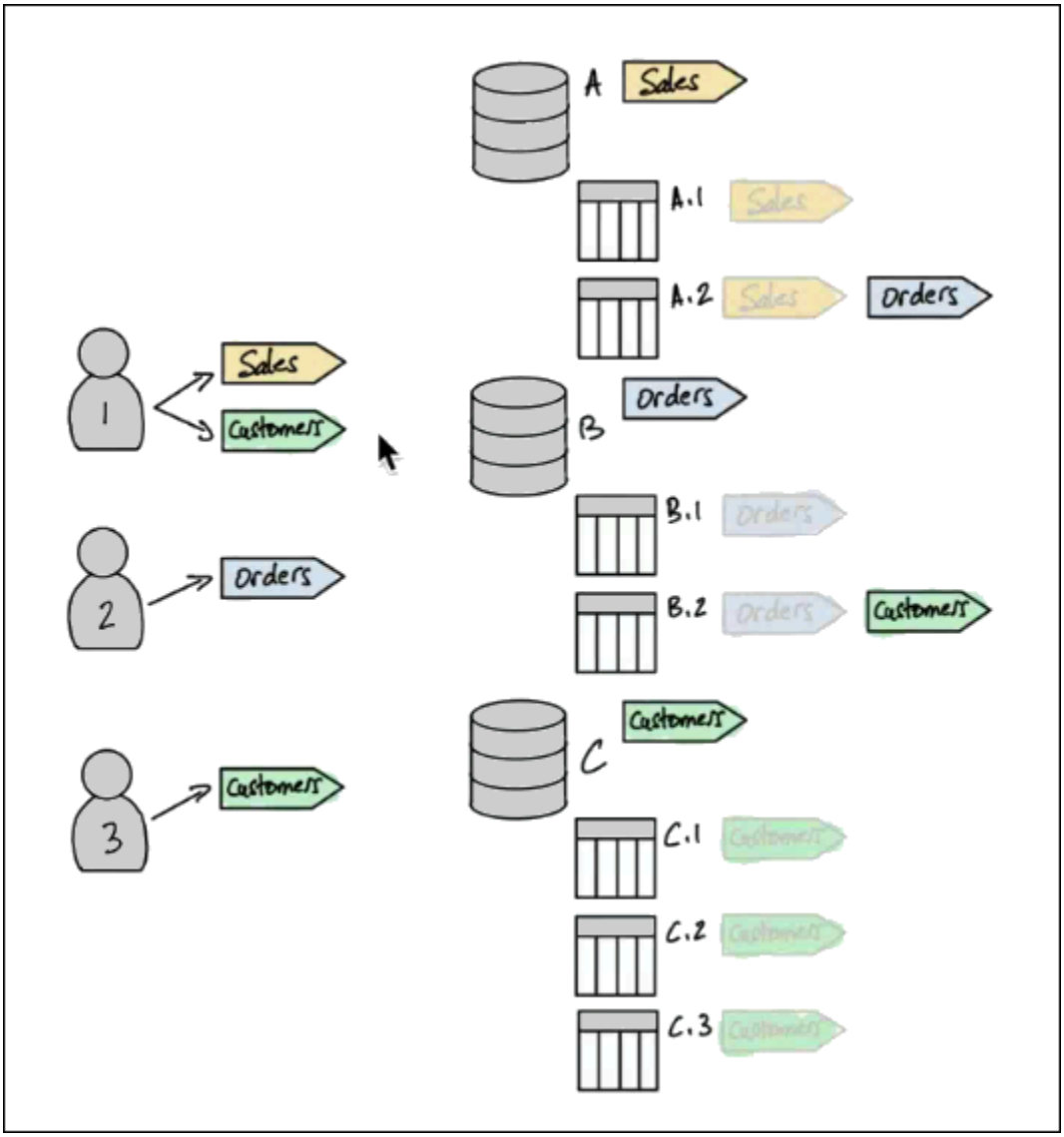
```
GRANT CREATE_TABLE ON Database A TO PRINCIPAL 1
GRANT SELECT, INSERT ON Table A.1 TO PRINCIPAL 1
GRANT SELECT, INSERT ON Table A.2 TO PRINCIPAL 1
GRANT SELECT, INSERT ON Table B.2 TO PRINCIPAL 1
...
GRANT SELECT, INSERT ON Table A.2 TO PRINCIPAL 2
GRANT CREATE_TABLE ON Database B TO PRINCIPAL 2
...
GRANT SELECT, INSERT ON Table C.3 TO PRINCIPAL 3
```

現在請考慮如何使用 LF-TBAC 授予許可。下圖指出您已將 LF 標籤指派給資料庫和資料表，並已將 LF 標籤的許可授予委託人。

在此範例中，LF 標籤代表資料湖的區域，其中包含企業資源規劃 (ERP) 應用程式套件不同模組的分析。您可以控制對各種模組分析資料的存取。所有 LF 標籤都有金鑰`module`和可能的值 `Sales`、`Orders`和 `Customers`。範例 LF-Tag 如下所示：

```
module=Sales
```

圖表僅顯示 LF 標籤值。



標記 Data Catalog 資源和繼承的指派

資料表從資料庫繼承 LF 標籤，資料欄從資料表繼承 LF 標籤。繼承的值可以覆寫。在上圖中，灰色的 LF 標籤會繼承。

由於繼承，資料湖管理員只需要對資源進行下列五個 LF 標籤指派（以虛擬程式碼表示）。

```
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
```

標記授予委託人

將 LF 標籤指派給資料庫和資料表之後，資料湖管理員必須僅對主體授予四個 LF 標籤，如下所示（在虛擬程式碼中）。

```
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
```

現在，具有 module=SalesLF-Tag 的委託人可以使用 LF-Tag module=Sales 存取 Data Catalog 資源（例如，資料庫 A)，具有 module=Customers LF-Tag 的委託人可以使用 LF-Tag module=Customers 存取資源，以此類推。

上述授予命令不完整。這是因為雖然它們透過 LF 標記資料目錄資源表示委託人具有許可，但它們不確切指出委託人在這些資源上擁有哪些 Lake Formation 許可（例如 SELECT、ALTER)。因此，以下虛擬程式碼命令更精確地表示如何透過 LF 標籤在 Data Catalog 資源上授予 Lake Formation 許可。

```
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
```

將它放在一起 - 產生資源的許可

鑑於上圖中指派給資料庫和資料表的 LF 標籤，以及圖中授予委託人的 LF 標籤，下表列出委託人在資料庫和資料表上擁有的 Lake Formation 許可。

Principal	透過 LF 標籤授予的許可
委託人 1	• CREATE_TABLE 資料庫 A 上的

Principal	透過 LF 標籤授予的許可
	• SELECT，在資料表 A.1 INSERT上 • SELECT，在資料表 B.2 INSERT上 • CREATE_TABLE 資料庫 C 上的 • SELECT，在資料表 C.1 INSERT上 • SELECT，在資料表 C.2 INSERT上 • SELECT，在資料表 C.3 INSERT上
委託人 2	• SELECT，在資料表 A.2 INSERT上 • CREATE_TABLE 資料庫 B 上的 • SELECT，在資料表 B.1 INSERT上
委託人 3	• SELECT，在資料表 B.2 INSERT上 • CREATE_TABLE 資料庫 C 上的 • SELECT，在資料表 C.1 INSERT上 • SELECT，在資料表 C.2 INSERT上 • SELECT，在資料表 C.3 INSERT上

底線

在此簡單範例中，使用五個指派操作和八個授予操作，資料湖管理員能夠指定 17 個許可。當有數十個資料庫和數百個資料表時，LF-TBAC 方法對具名資源方法的優勢會變得明確。在假設需要授予每個主體對每個資源的存取權的情況下，其中 n(P)是主體數目，而 n(R)是資源數目：

- 使用具名資源方法時，所需的授予數量為 n(P) × n(R)。
- 透過 LF-TBAC 方法，使用單一 LF-Tag，授予委託人和指派資源的總數為 n(P) + n(R)。

另請參閱

- 管理中繼資料存取控制的 LF 標籤
- 使用 LF-TBAC 方法授予資料湖許可

主題

- 管理中繼資料存取控制的 LF 標籤
- 管理中繼資料存取控制的 LF-Tag 表達式
- 管理 LF-Tag 值許可

管理中繼資料存取控制的 LF 標籤

若要使用 Lake Formation 標籤型存取控制 (LF-TBAC) 方法來保護 Data Catalog 物件，例如目錄、資料庫、資料表、檢視和資料欄，您可以建立 LF 標籤、將它們指派給資源，並將 LF 標籤許可授予委託人。

您必須先定義 LF 標籤，才能將 LF 標籤指派給 Data Catalog 物件或將許可授予委託人。只有具有 LF 標籤建立者許可的資料湖管理員或主體才能建立 LF 標籤。

LF 標籤建立者

LF-Tag 建立者是非管理員主體，具有建立和管理 LF-Tags 的許可。資料湖管理員可以使用 Lake Formation 主控台或 CLI 新增 LF 標籤建立者。LF-Tag 建立者具有隱含 Lake Formation 許可來更新和刪除 LF-Tags、將 LF-Tags 指派給資源，以及將 LF-Tag 許可和 LF-Tag 值許可授予其他主體。

透過 LF 標籤建立者角色，資料湖管理員可以將標籤管理任務委派給非管理員主體，例如建立和更新標籤金鑰和值。資料湖管理員也可以授予 LF 標籤建立者可授予的`Create LF-Tag`許可。然後，LF-Tag 建立者可以將建立 LF-Tag 的許可授予其他主體。

您可以在 LF 標籤上授予兩種類型的許可：

- LF 標籤許可 - `Alter`、 `Create LF-Tag`和 `Drop`。建立、更新和刪除 LF 標籤需要這些許可。

  資料湖管理員和 LF 標籤建立者會隱含地對其建立的 LF 標籤擁有這些許可，並且可以明確地將這些許可授予委託人，以管理資料湖中的標籤。
- LF 標籤鍵/值對許可 - `Describe`、 `Assign`和 `Grant with LF-Tag expressions`。將 LF 標籤指派給 Data Catalog 物件，以及使用 Lake Formation 標籤型存取控制將資源的許可授予委託人時，需要這些許可。LF-Tag 建立者在建立 LF-Tags 時隱含地接收這些許可。

收到`Create LF-Tag`許可並成功建立 LF 標籤後，LF 標籤建立者可以將 LF 標籤指派給資源，並將 LF 標籤許可 (`Create LF-Tag`、`Drop`、 `Alter`和) 授予其他非管理主體，以管理資料湖中的標籤。您可以使用 Lake Formation 主控台、API 或 AWS Command Line Interface () 來管理 LF 標籤AWS CLI。

> **Note**
>
> 資料湖管理員具有隱含 Lake Formation 許可，可建立、更新和刪除 LF 標籤、將 LF 標籤指派給資源，以及將 LF 標籤許可授予委託人。

如需最佳實務和考量事項，請參閱 Lake Formation 標籤型存取控制最佳實務和考量事項

主題

- 新增 LF 標籤建立者
- 建立 LF 標籤
- 更新 LF 標籤
- 刪除 LF 標籤
- 列出 LF 標籤
- 將 LF 標籤指派給 Data Catalog 資源
- 檢視指派給資源的 LF 標籤
- 檢視 LF-Tag 指派給的資源
- LF 標籤的生命週期
- Lake Formation 標籤型存取控制與 IAM 屬性型存取控制的比較

> **另請參閱**
>
> - 管理 LF-Tag 值許可
> - 使用 LF-TBAC 方法授予資料湖許可
> - Lake Formation 標籤型存取控制

新增 LF 標籤建立者

根據預設，資料湖管理員可以建立、更新和刪除 LF 標籤、將標籤指派給 Data Catalog 物件，以及將標籤許可授予委託人。如果您想要將標籤建立和管理操作委派給非管理員主體，資料湖管理員可以建立 LF 標籤建立者角色，並將 Lake Formation `Create LF-Tag`許可授予角色。透過可授予的`Create LF-Tag`許可，LF-Tag 建立者可以將標籤建立和維護任務委派給其他非管理主體。

若要讓資料湖管理員將 LF-Tags 指派給 Data Catalog 資源，他們需要授予自己非由其建立的 LF-Tags 的關聯許可。

> **Note**
>
> 跨帳戶許可授予只能包含 `Describe`和 `Associate`許可。您無法將 `Create LF-Tag`、`Alter`、 `Drop`和 `Grant with LFTag expressions`許可授予不同帳戶中的委託人。

主題

- 建立 LF 標籤所需的 IAM 許可
- 新增 LF 標籤建立者

> **另請參閱**
>
> - 管理 LF-Tag 值許可
> - 使用 LF-TBAC 方法授予資料湖許可
> - Lake Formation 標籤型存取控制

建立 LF 標籤所需的 IAM 許可

您必須設定許可，以允許 Lake Formation 主體建立 LF 標籤。將下列陳述式新增至需要建立 LF 標籤的主體的許可政策。

> **Note**
>
> 雖然資料湖管理員具有隱含 Lake Formation 許可來建立、更新和刪除 LF 標籤、將 LF 標籤指派給資源，以及將 LF 標籤授予委託人，但資料湖管理員也需要下列 IAM 許可。

如需詳細資訊，請參閱Lake Formation 角色和 IAM 許可參考。

```
{
"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"
     ]
 }
```

將 LF-Tags 指派給資源並將 LF-Tags 授予委託人的委託人必須具有相同的許可，但 `CreateLFTag`、`UpdateLFTag`和 `DeleteLFTag`許可除外。

新增 LF 標籤建立者

LF 標籤建立者可以建立 LF 標籤、更新標籤索引鍵和值、刪除標籤、將標籤與 Data Catalog 資源建立關聯，以及使用 LF-TBAC 方法將 Data Catalog 資源的許可授予委託人。LF-Tag 建立者也可以將這些許可授予委託人。

您可以使用 AWS Lake Formation 主控台、 API 或 AWS Command Line Interface () 來建立 LF 標籤建立者角色AWS CLI。

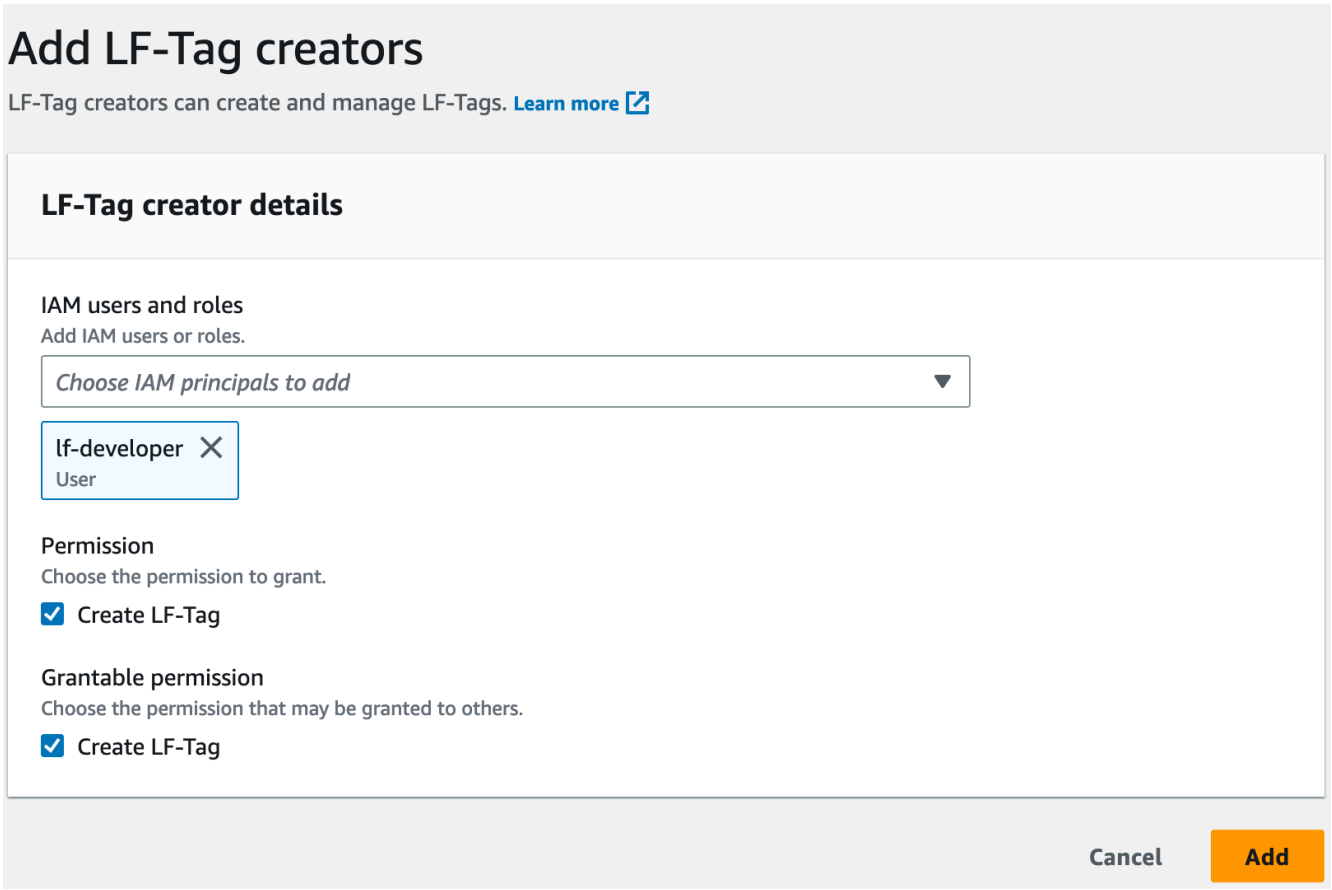
console

新增 LF 標籤建立者

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

   以 Datalake 管理員身分登入。
2. 在導覽窗格的許可下，選擇 LF 標籤和許可。

   在 LF 標籤和許可頁面上，選擇 LF 標籤建立者區段，然後選擇新增 LF 標籤建立者。



3. 在新增 LF 標籤建立者頁面上，選擇具有建立 LF 標籤所需許可的 IAM 角色或使用者。
4. 啟用`Create LF-Tag`許可核取方塊。
5. （選用）若要讓選取的委託人將`Create LF-Tag`許可授予委託人，請選擇可授予`Create LF-Tag`許可。
6. 選擇新增。

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"
    ]
}
```

以下是 LF 標籤建立者角色可用的許可：

權限	描述
Drop	在 LF-Tag 具有此許可的主體可以從資料湖刪除 LF-Tag。主體會取得 LF-Tag 資源所有標籤值的隱含Describe許可。
Alter	在 LF 標籤具有此許可的委託人可以從 LF 標籤新增或移除標籤值。委託人會取得 LF 標籤之所有標籤值的隱含Alter許可。
Describe	在 LF-Tag 上具有此許可的委託人可以在將 LF-Tags 指派給資源或授予 LF-Tags 許可時，檢視 LF-Tag 及其值。您可以在Describe所有索引鍵值或特定值上授予 。
Associate	在 LF-Tag 上具有此許可的主體可以將 LF-Tag 指派給 Data Catalog 資源。隱Associate 含授予 Describe。
Grant with LF-Tag expression	在 LF-Tag 具有此許可的主體可以使用 LF-Tag 金鑰和值授予資料目錄資源的許可。隱Grant with LF-Tag expression 含授予 Describe。

這些許可是可授予的。已使用授予選項授予這些許可的委託人可以將其授予其他委託人。

建立 LF 標籤

所有 LF 標籤都必須在 Lake Formation 中定義，才能使用。LF 標籤包含金鑰，以及金鑰的一或多個可能值。

資料湖管理員為 LF-Tag 建立者角色設定必要的 IAM 許可和 Lake Formation 許可後，主體可以建立 LF-Tag。LF-Tag 建立者會取得隱含許可，以更新或移除 LF-Tag 中的任何標籤值，並刪除 LF-Tag。

您可以使用 AWS Lake Formation 主控台、 API 或 () 來建立 AWS Command Line Interface LF 標籤 AWS CLI。

Console

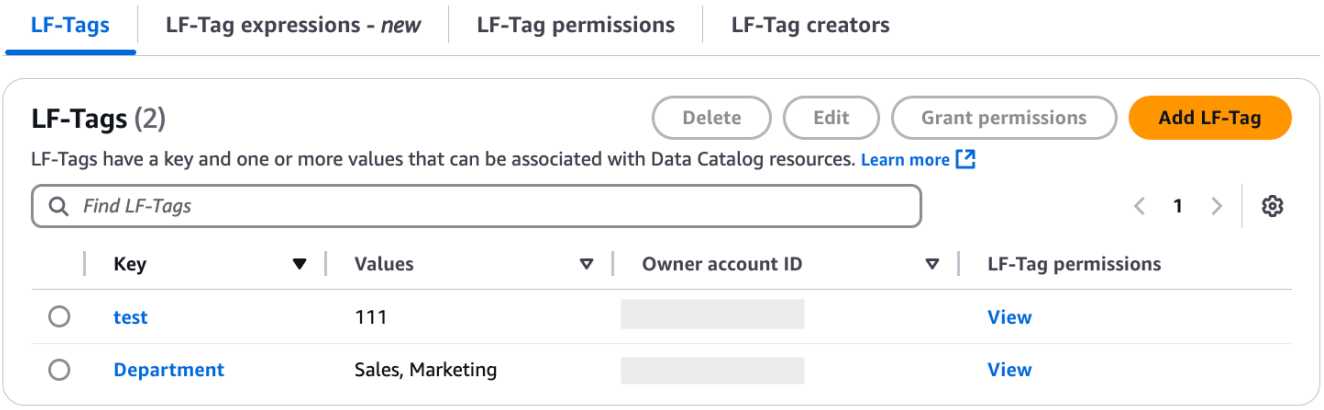
建立 LF 標籤

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

   以具有 LF-Tag 建立者許可的主體身分登入，或以資料湖管理員身分登入。

2. 在導覽窗格的許可、LF 標籤和許可下，選擇 LF 標籤。

   LF-Tags 頁面隨即出現。



3. 選擇新增 LF 標籤。
4. 在新增 LF 標籤對話方塊中，輸入金鑰和一或多個值。

   每個金鑰必須至少有一個值。若要輸入多個值，請輸入逗號分隔清單，然後按 Enter，或一次輸入一個值，然後在每個值之後選擇新增。允許的值數目上限為 1000。

5. 選擇 Add tag (新增標籤)。

AWS CLI

建立 LF 標籤

- 輸入`create-lf-tag`命令。

  下列範例會建立具有索引鍵`module`和值 和 `Customers`的 LF 標籤`Orders`。

```
aws lakeformation create-lf-tag --tag-key module --tag-values Customers Orders
```

身為標籤建立者 ，委託人會取得此 LF 標籤的Alter許可，並且可以從此 LF 標籤更新或移除任何標籤值。LF-Tag 建立者主體也可以授予Alter許可給另一個主體，以更新和移除此 LF-Tag 上的標籤值。

更新 LF 標籤

您可以透過新增或刪除允許的金鑰值來更新您擁有 Alter許可的 LF 標籤。您無法變更 LF-Tag 金鑰。若要變更金鑰，請刪除 LF 標籤，並使用必要的金鑰新增標籤。除了 Alter 許可之外，您也需要 lakeformation:UpdateLFTag IAM 許可才能更新值。

當您刪除 LF-Tag 值時，不會對任何 Data Catalog 資源上是否存在該 LF-Tag 值執行檢查。如果已刪除的 LF-Tag 值與資源相關聯，則不再顯示該資源，且在該鍵值對上授予許可的任何主體都不再具有許可。

刪除 LF-Tag 值之前，您可以選擇使用 remove-lf-tags-from-resource 命令從具有要刪除值的資料目錄資源中移除 LF-Tag，然後使用您要保留的值重新標記資源。

只有擁有 LF-Tag Alter許可的資料湖管理員、LF-Tag 建立者和主體才能更新 LF-Tag。

您可以使用 AWS Lake Formation 主控台、 API 或 () 來更新 AWS Command Line Interface LF 標籤 AWS CLI。

Console

更新 LF 標籤 （主控台）

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

   以資料湖管理員、LF 標籤建立者或具有 LF 標籤Alter許可的主體身分登入。
2. 在導覽窗格的許可、LF 標籤和許可下，選擇 LF 標籤。
3. 在 LF-Tags 頁面上，選取 LF-Tag，然後選擇編輯。
4. 在編輯 LF-Tag 對話方塊中，新增或移除 LF-Tag 值。

   若要新增多個值，請在值欄位中，輸入逗號分隔清單並按 Enter，或一次輸入一個值，或在每個值之後選擇新增。
5. 選擇儲存。

AWS CLI

更新 LF 標籤 (AWS CLI)

- 輸入 `update-lf-tag`命令。提供下列其中一個或兩個引數：
  - `--tag-values-to-add`
  - `--tag-values-to-delete`

Example

下列範例會將 值取代`vp`為 `vice-president` LF-Tag 金鑰 的值`level`。

```
aws lakeformation update-lf-tag --tag-key level --tag-values-to-add vice-president
--tag-values-to-delete vp
```

刪除 LF 標籤

您可以刪除不再使用的 LF 標籤。不會對 Data Catalog 資源上是否存在 LF 標籤執行檢查。如果已刪除的 LF-Tag 與資源相關聯，則不再顯示該資源，且在該 LF-Tag 上授予許可的任何主體都不再具有許可。

在刪除 LF-Tag 之前，您可以選擇使用 `remove-lf-tags-from-resource`命令從所有資源中移除 LF-Tag。

只有資料湖管理員、LF 標籤建立者或具有 LF 標籤`Drop`許可的主體可以刪除 LF 標籤。除了 `Drop`許可之外，委託人還需要 `lakeformation:DeleteLFTag` IAM 許可才能刪除 LF 標籤。

您可以使用 AWS Lake Formation 主控台、 API 或 () 刪除 AWS Command Line Interface LF 標籤 AWS CLI。

Console

刪除 LF 標籤 （主控台）

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

   以資料湖管理員身分登入。
2. 在導覽窗格中的許可、LF 標籤和許可下，選擇 LF 標籤。
3. 在 LF-Tags 頁面上，選取 LF-Tag，然後選擇刪除。

4. 在刪除標籤環境？對話方塊中，若要確認刪除，請在指定欄位中輸入 LF-Tag 金鑰值，然後選擇刪除。

AWS CLI

刪除 LF 標籤 (AWS CLI)

- 輸入`delete-lf-tag`命令。提供要刪除之 LF-Tag 的金鑰。

  Example

  下列範例使用金鑰 刪除 LF-Tag`region`。

```
aws lakeformation delete-lf-tag --tag-key region
```

列出 LF 標籤

您可以列出您擁有 `Describe`或 `Associate`許可的 LF 標籤。每個 LF-Tag 金鑰列出的值是您擁有許可的值。

LF-Tag 建立者具有隱含許可，可查看他們已建立的 LF-Tag。

資料湖管理員可以看到本機 AWS 帳戶中定義的所有 LF 標籤，以及已從外部帳戶將 `Describe`和 `Associate`許可授予本機帳戶的所有 LF 標籤。資料湖管理員可以看到所有 LF 標籤的所有值。

您可以使用 AWS Lake Formation 主控台、 API 或 () AWS Command Line Interface 列出 LF 標籤 AWS CLI。

Console

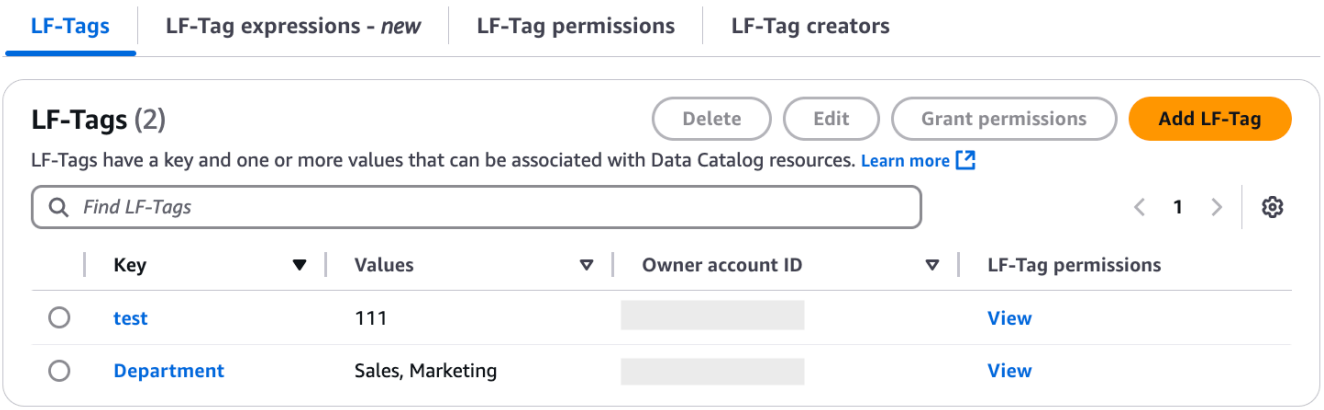
列出 LF 標籤 （主控台）

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

   以 LF-Tag 建立者、資料湖管理員或已授予 LF-Tags 許可且具有 IAM `lakeformation:ListLFTags` 許可的主體身分登入。

2. 在導覽窗格的許可、LF 標籤和許可下，選擇 LF 標籤。

   LF-Tags 頁面隨即出現。



檢查擁有者帳戶 ID 欄，以判斷從外部帳戶與您的帳戶共用的 LF 標籤。

AWS CLI

列出 LF 標籤 (AWS CLI)

- 以資料湖管理員或已授予 LF-Tags 許可且具有 IAM `lakeformation:ListLFTags` 許可的委託人身分執行下列命令。

```
aws lakeformation list-lf-tags
```

輸出類似如下。

```
{
    "LFTags": [
        {
            "CatalogId": "111122223333",
            "TagKey": "level",
            "TagValues": [
                "director",
                "vp",
                "c-level"
            ]
        },
        {
            "CatalogId": "111122223333",
            "TagKey": "module",
            "TagValues": [
                "Orders",
                "Sales",
```

```
                "Customers"
            ]
        }
    ]
}
```

若要查看從外部帳戶授予的 LF 標籤，請包含命令選項 `--resource-share-type ALL`。

```
aws lakeformation list-lf-tags --resource-share-type ALL
```

輸出類似如下。請記下 `NextToken`金鑰，這表示要列出更多項目。

```
{
    "LFTags": [
        {
            "CatalogId": "111122223333",
            "TagKey": "level",
            "TagValues": [
                "director",
                "vp",
                "c-level"
            ]
        },
        {
            "CatalogId": "111122223333",
            "TagKey": "module",
            "TagValues": [
                "Orders",
                "Sales",
                "Customers"
            ]
        }
    ],
    "NextToken": "eyJleHBpcmF0aW...ZXh0Ijp0cnVlfQ=="
}
```

重複 命令，並新增 `--next-token`引數，以檢視從外部帳戶授予的任何剩餘本機 LF 標籤和 LF 標籤。來自外部帳戶的 LF 標籤一律位於單獨的頁面上。

```
aws lakeformation list-lf-tags --resource-share-type ALL
--next-token eyJleHBpcmF0aW...ZXh0Ijp0cnVlfQ==
```

```
{
    "LFTags": [
        {
            "CatalogId": "123456789012",
            "TagKey": "region",
            "TagValues": [
                "central",
                "south"
            ]
        }
    ]
}
```

API

您可以使用 Lake Formation SDKs，列出申請者有權檢視的標籤。

```
import boto3

client = boto3.client('lakeformation')
...

response = client.list_lf_tags(
    CatalogId='string',
    ResourceShareType='ALL',
    MaxResults=50'
)
```

此命令會傳回具有下列結構的dict物件：

```
{
    'LFTags': [
        {
            'CatalogId': 'string',
            'TagKey': 'string',
            'TagValues': [
                'string',
            ]
        },
    ],
    'NextToken': 'string'
```

```
}
```

如需所需許可的詳細資訊，請參閱Lake Formation 角色和 IAM 許可參考。

將 LF 標籤指派給 Data Catalog 資源

您可以將 LF 標籤指派給 Data Catalog 資源（資料庫、資料表和欄），以控制對這些資源的存取。只有獲授予相符 LF 標籤的主體（以及獲授予具名資源方法存取權的主體）才能存取資源。

如果資料表從資料庫繼承 LF 標籤，或資料欄從資料表繼承 LF 標籤，您可以透過將新值指派給 LF 標籤索引鍵來覆寫繼承的值。

您可以指派給資源的 LF 標籤數目上限為 50。

主題

- 管理指派給資源的標籤的需求
- 將 LF 標籤指派給資料表資料欄
- 將 LF 標籤指派給 Data Catalog 資源
- 更新資源的 LF 標籤
- 從資源移除 LF-Tag

管理指派給資源的標籤的需求

若要將 LF 標籤指派給 Data Catalog 資源，您必須：

- 擁有 LF-Tag 上的 Lake Formation ASSOCIATE許可。
- 擁有 IAM `lakeformation:AddLFTagsToResource`許可。
- 在 Glue 資料庫上擁有 glue：GetDatabase 許可。
- 成為資源擁有者（建立者）、對具有 GRANT選項的資源具有 Super Lake Formation 許可，或具有具有 GRANT選項的下列許可：
  - 對於相同 AWS 帳戶中的資料庫：DESCRIBE、ALTER、 CREATE_TABLE和 DROP
  - 對於外部帳戶中的資料庫： DESCRIBECREATE_TABLE和 ALTER
  - 對於資料表（和資料欄）：DESCRIBE、ALTER、DROPINSERT、SELECT、 和 DELETE

此外，LF-Tag 及其指派的資源必須位於相同的 AWS 帳戶中。

若要從 Data Catalog 資源中移除 LF-Tag，您必須符合這些要求，並具有 `lakeformation:RemoveLFTagsFromResource` IAM 許可。

將 LF 標籤指派給資料表資料欄

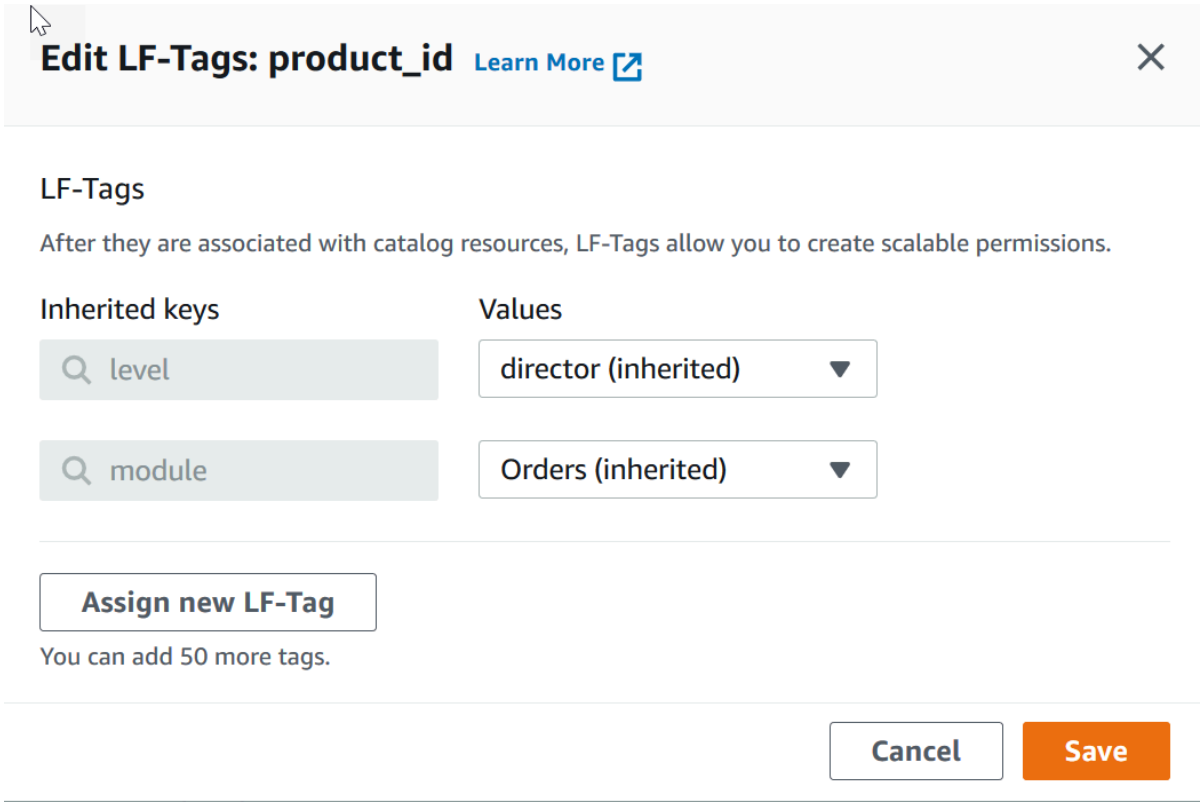
將 LF 標籤指派給資料表欄（主控台）

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

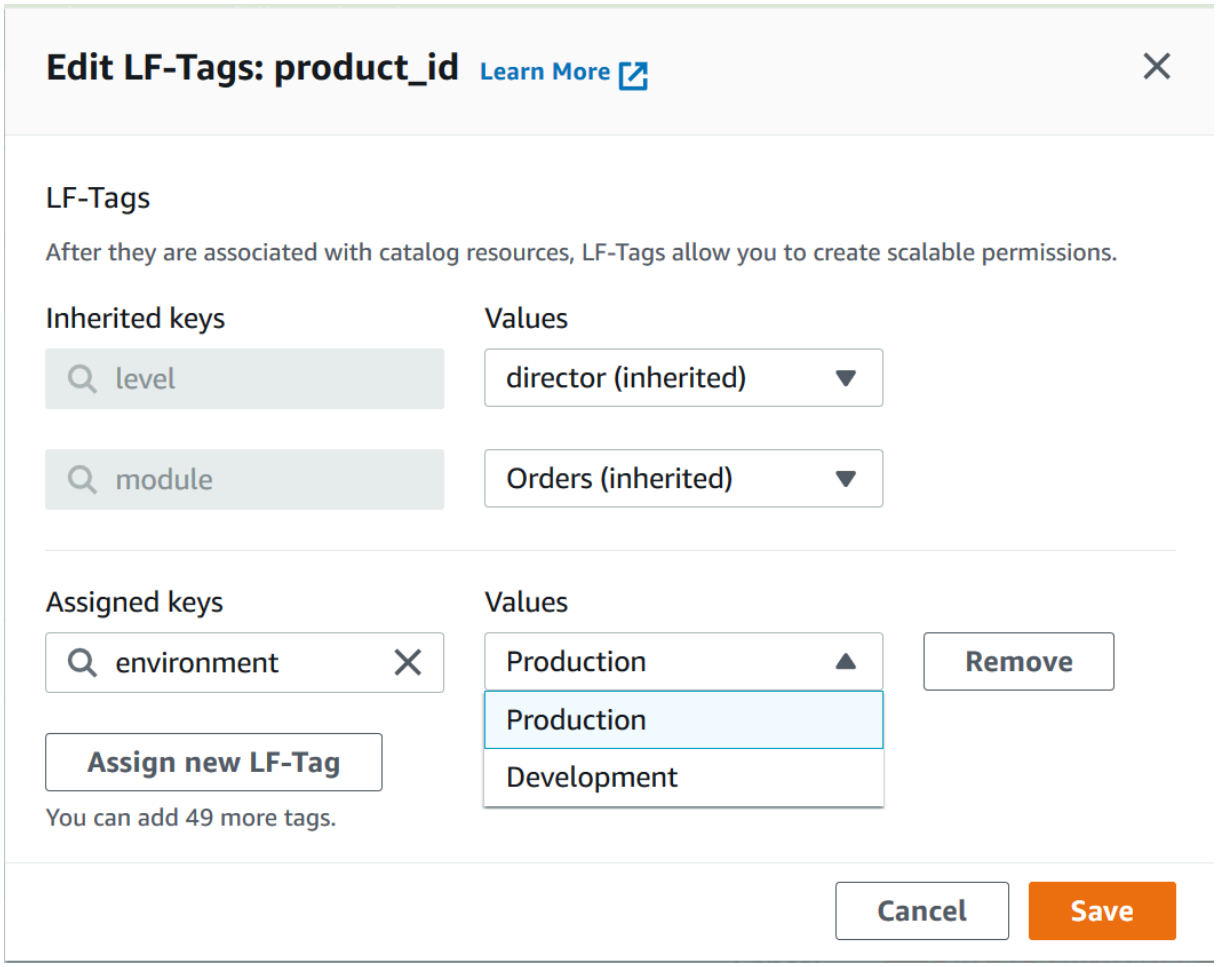
   以符合上述要求的使用者身分登入。
2. 在導覽窗格中，選擇 Tables (資料表)。
3. 選擇資料表名稱（而非資料表名稱旁的選項按鈕）。
4. 在資料表詳細資訊頁面的結構描述區段中，選擇編輯結構描述。
5. 在編輯結構描述頁面上，選取一或多個資料欄，然後選擇編輯 LF 標籤。

   > **Note**
   >
   > 如果您想要新增或刪除資料欄並儲存新版本，請先執行此操作。然後編輯 LF 標籤。

   編輯 LF 標籤對話方塊隨即出現，並顯示從資料表繼承的任何 LF 標籤。



6. （選用）對於繼承索引鍵欄位旁的值清單，選擇值以覆寫繼承的值。
7. （選用）選擇指派新的 LF 標籤。然後，對於指派的金鑰，選擇金鑰，對於值，選擇金鑰的值。



8. （選用）再次選擇指派新的 LF 標籤以新增另一個 LF 標籤。
9. 選擇儲存。

將 LF 標籤指派給 Data Catalog 資源

Console

將 LF 標籤指派給 Data Catalog 資料庫或資料表

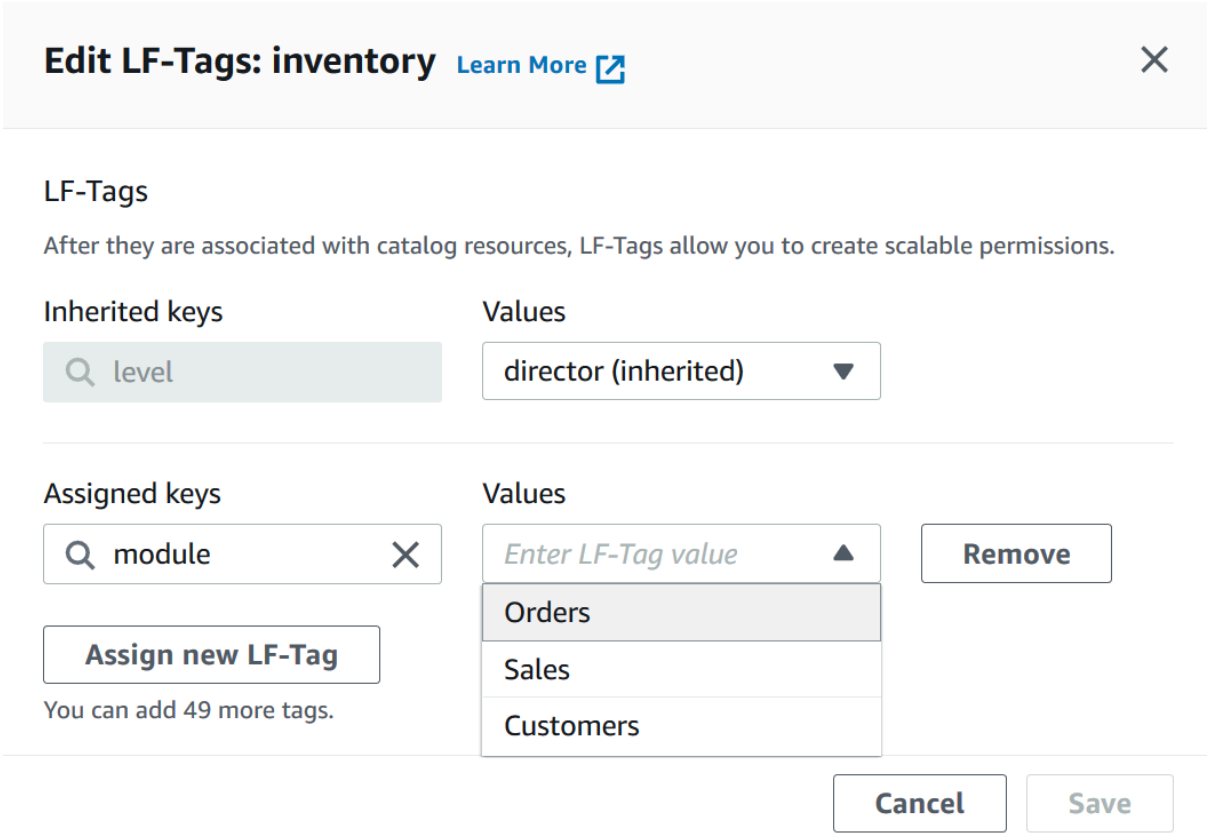
1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

   以符合先前所列要求的使用者身分登入。
2. 在導覽窗格的資料目錄下，執行下列其中一項操作：
   - 若要將 LF 標籤指派給資料庫，請選擇資料庫。
   - 若要將 LF 標籤指派給資料表，請選擇資料表。

3. 選擇資料庫或資料表，然後在動作功能表上，選擇編輯 LF 標籤。

   編輯 LF-Tags： ***resource-name*** 對話方塊隨即出現。

   如果資料表繼承包含資料庫的 LF 標籤，則視窗會顯示繼承的 LF 標籤。否則，它會顯示文字「沒有與資源相關聯的繼承 LF 標籤」。



4. （選用）如果資料表繼承了 LF 標籤，對於繼承索引鍵欄位旁的值清單，您可以選擇值來覆寫繼承的值。
5. 若要指派新的 LF 標籤，請執行下列步驟：

   a. 選擇指派新的 LF 標籤。

   b. 在指派的金鑰欄位中，選擇 LF-Tag 金鑰，然後在值欄位中，選擇值。

   c. （選用）再次選擇指派新的 LF 標籤，以指派額外的 LF 標籤。
6. 選擇儲存。

AWS CLI

將 LF 標籤指派給 Data Catalog 資源

- 執行 add-lf-tags-to-resource 命令。

  下列範例會將 LF-Tag module=orders 指派給資料庫 orders中的資料表erp。它使用 --lf-tags引數的捷徑語法。的 CatalogID 屬性--lf-tags是選用的。如果未提供，則會假設資源的目錄 ID （在此情況下為 資料表）。

```
aws lakeformation add-lf-tags-to-resource --resource '{ "Table":
 {"DatabaseName":"erp", "Name":"orders"}}' --lf-tags
 CatalogId=111122223333,TagKey=module,TagValues=orders
```

  以下是命令成功時的輸出。

```
{
    "Failures": []
}
```

  下一個範例會將兩個 LF 標籤指派給sales資料表，並使用--lf-tags引數的 JSON 語法。

```
aws lakeformation add-lf-tags-to-resource --resource '{ "Table":
 {"DatabaseName":"erp", "Name":"sales"}}' --lf-tags '[{"TagKey":
 "module","TagValues": ["sales"]},{"TagKey": "environment","TagValues":
 ["development"]}]'
```

  下一個範例會將 LF-Tag level=director 指派給資料表 的資料total欄sales。

```
aws lakeformation add-lf-tags-to-resource --resource '{ "TableWithColumns":
 {"DatabaseName":"erp", "Name":"sales", "ColumnNames":["total"]}}' --lf-tags
 TagKey=level,TagValues=director
```

更新資源的 LF 標籤

更新 Data Catalog 資源的 LF 標籤 (AWS CLI)

- 使用 add-lf-tags-to-resource命令，如先前程序所述。

新增具有與現有 LF 標籤相同金鑰的 LF 標籤，但具有不同值的 LF 標籤會更新現有值。

從資源移除 LF-Tag

移除 Data Catalog 資源的 LF 標籤 (AWS CLI)

- 執行 `remove-lf-tags-from-resource` 命令。

  如果資料表的 LF-Tag 值會覆寫從父資料庫繼承的值，則從資料表中移除該 LF-Tag 會還原繼承的值。此行為也適用於覆寫從資料表繼承之索引鍵值的資料欄。

  下列範例會從`sales`資料表的 資料`total`欄移除 `level=director` LF 標籤。的 `CatalogID` 屬性`--lf-tags`是選用的。如果未提供，則會假設資源的目錄 ID （在此情況下為 資料表）。

```
aws lakeformation remove-lf-tags-from-resource
--resource ' { "TableWithColumns":
{ "DatabaseName": "erp",  "Name": "sales",  "ColumnNames":[ "total"]}}'
--lf-tags  CatalogId=111122223333,TagKey=level,TagValues=director
```

檢視指派給資源的 LF 標籤

您可以檢視指派給 Data Catalog 資源的 LF 標籤。您必須在 LF 標籤上擁有 `DESCRIBE`或 `ASSOCIATE`許可才能檢視它。

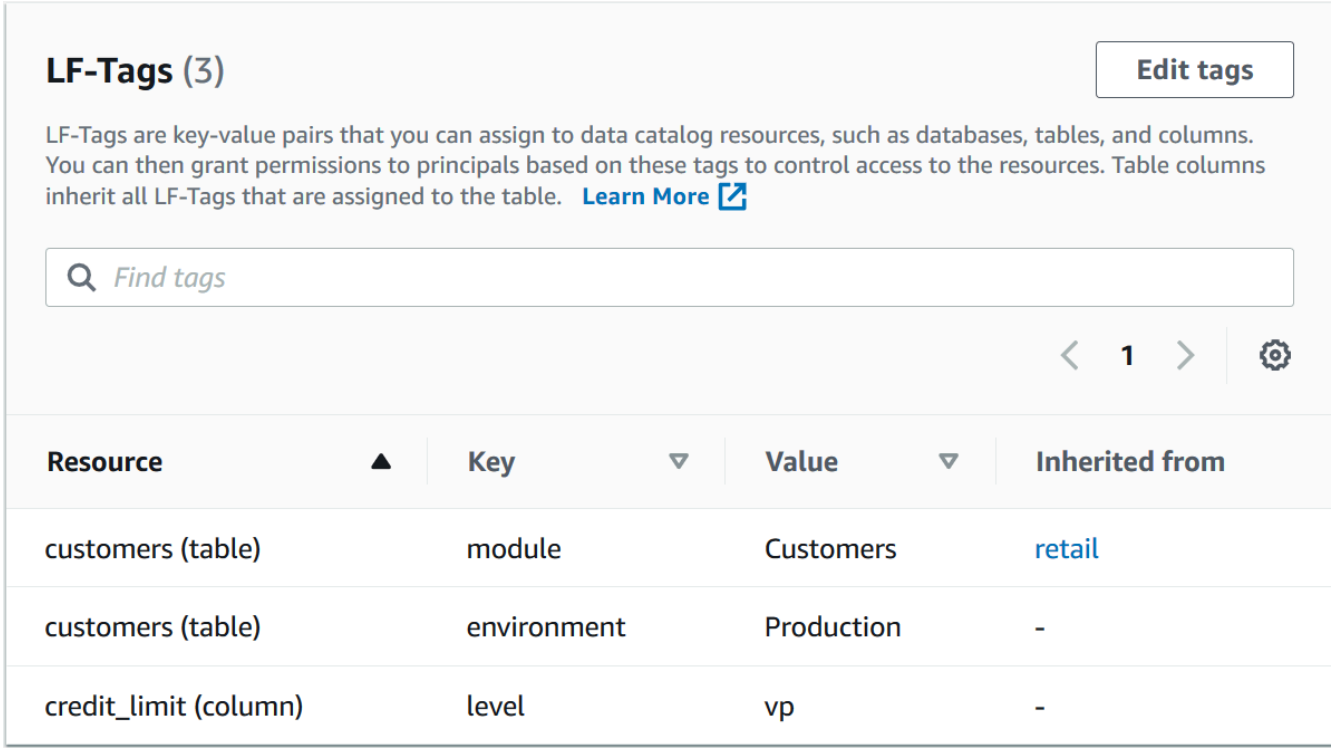
Console

檢視指派給資源的 LF 標籤 （主控台）

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

   以資料湖管理員、資源擁有者或已授予資源 Lake Formation 許可的使用者身分登入。
2. 在導覽窗格的資料目錄標題下，執行下列其中一項操作：
   - 若要檢視指派給資料庫的 LF 標籤，請選擇資料庫。
   - 若要檢視指派給資料表的 LF 標籤，請選擇資料表。
3. 在資料表或資料庫頁面上，選擇資料庫或資料表的名稱。然後在詳細資訊頁面上，向下捲動至 LF 標籤區段。

下列螢幕擷取畫面顯示指派給customers資料表的 LF 標籤，其中包含在retail資料庫中。module LF 標籤繼承自資料庫。資料credit_limit欄已指派 level=vp LF 標籤。



AWS CLI

檢視指派給資源的 LF 標籤 (AWS CLI)

- 輸入與以下相似的命令。

```
aws lakeformation get-resource-lf-tags --show-assigned-lf-tags --
resource '{ "Table": {"CatalogId":"111122223333", "DatabaseName":"erp",
 "Name":"sales"}}'
```

命令會傳回下列輸出：

```
{
    "TableTags": [
        {
            "CatalogId": "111122223333",
            "TagKey": "module",
            "TagValues": [
                "sales"
```

```
            ]
        },
        {
            "CatalogId": "111122223333",
            "TagKey": "environment",
            "TagValues": [
                "development"
            ]
        }
    ],
    "ColumnTags": [
        {
            "Name": "total",
            "Tags": [
                {
                    "CatalogId": "111122223333",
                    "TagKey": "level",
                    "TagValues": [
                        "director"
                    ]
                }
            ]
        }
    ]
}
```

此輸出只會顯示明確指派而非繼承的 LF 標籤。如果您想要查看所有資料欄上的所有 LF 標籤，包括繼承的 LF 標籤，請省略 `--show-assigned-lf-tags`選項。

檢視 LF-Tag 指派給的資源

您可以檢視指定給特定 LF-Tag 金鑰的所有 Data Catalog 資源。若要這樣做，您需要下列 Lake Formation 許可：

- `Describe` 或 `Associate` 的 LF 標籤。
- `Describe` 或任何其他資源的 Lake Formation 許可。

此外，您需要下列 AWS Identity and Access Management (IAM) 許可：

- `lakeformation:SearchDatabasesByLFTags`
- `lakeformation:SearchTablesByLFTags`

Console

檢視 LF-Tag 指派給的資源 （主控台）

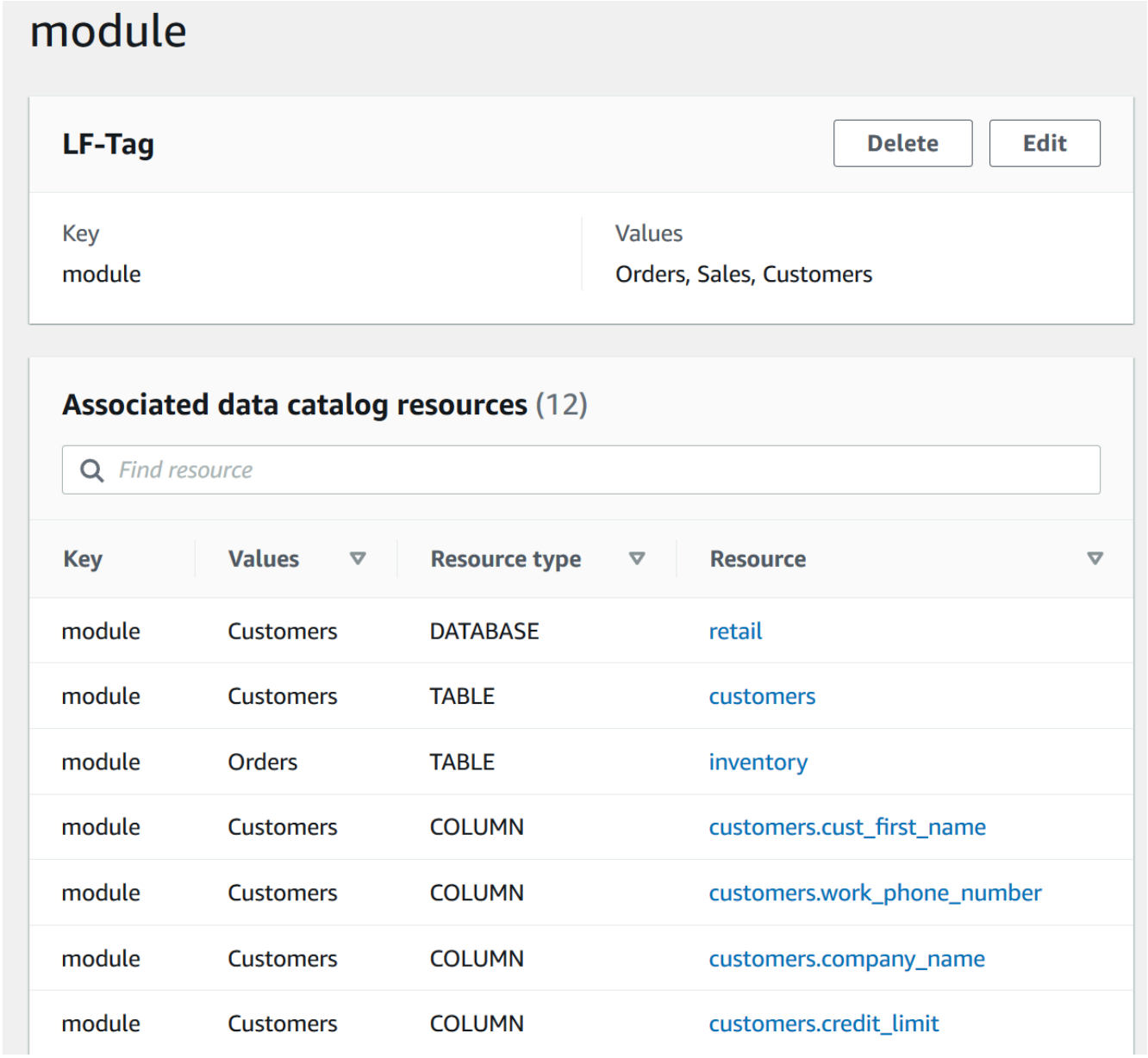
1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

   以資料湖管理員或符合先前列出的要求的使用者身分登入。

2. 在導覽窗格的許可、LF 標籤和許可下，選擇 LF 標籤。

3. 選擇 LF-Tag 金鑰 （而非金鑰名稱旁的選項按鈕）。

   LF-Tag 詳細資訊頁面會顯示已指派 LF-Tag 的資源清單。



AWS CLI

檢視 LF-Tag 指派給的資源

- 執行 `search-tables-by-lf-tags`或 `search-databases-by-lf-tags`命令。

  Example

  下列範例列出已指派 `level=vp` LF 標籤的資料表和資料欄。對於列出的每個資料表和資料欄，會輸出資料表或資料欄的所有指派 LF 標籤，而不只是搜尋表達式。

```
aws lakeformation search-tables-by-lf-tags --expression
 TagKey=level,TagValues=vp
```

如需所需許可的詳細資訊，請參閱Lake Formation 角色和 IAM 許可參考。

LF 標籤的生命週期

1. LF 標籤建立者 Michael 建立 LF 標籤 `module=Customers`。
2. Michael 在 LF 標籤`Associate`上授予資料工程師 Eduardo。`Associate` 隱含授予 `Describe`。
3. Michael 使用授予選項將資料表`Super`上的 Michael 授予 `Custs` Eduardo，以便 Eduardo 可以將 LF 標籤指派給資料表。如需詳細資訊，請參閱將 LF 標籤指派給 Data Catalog 資源。
4. Eduardo 會將 LF 標籤指派給`module=customers`資料表 `Custs`。
5. Michael 向資料工程師 Sandra （虛擬程式碼） 授予下列授權。

```
GRANT (SELECT, INSERT ON TABLES) ON TAGS module=customers TO Sandra WITH GRANT OPTION
```

6. Sandra 授予資料分析師 Maria 下列許可。

```
GRANT (SELECT ON TABLES) ON TAGS module=customers TO Maria
```

Maria 現在可以在`Custs`資料表上執行查詢。

另請參閱

- 中繼資料存取控制

Lake Formation 標籤型存取控制與 IAM 屬性型存取控制的比較

屬性型存取控制 (ABAC) 是一種授權策略，可根據屬性來定義許可。在 中 AWS，這些屬性稱為標籤。您可以將標籤連接至 IAM 資源，包括 IAM 實體（使用者或角色）和 AWS 資源。您可以為您的 IAM 主體建立單一 ABAC 政策或是一組政策。這些 ABAC 政策可以設計成在主體的標籤與資源標籤相符時允許操作。ABAC 在成長快速的環境中相當有幫助，並能在政策管理變得繁瑣時提供協助。

雲端安全與控管團隊使用 IAM 來定義所有資源的存取政策和安全許可，包括 Amazon S3 儲存貯體、Amazon EC2 執行個體，以及您可以使用 ARN 參考的任何資源。IAM 政策會定義資料湖資源的廣泛（粗細）許可，例如，允許或拒絕 Amazon S3 儲存貯體或字首層級或資料庫層級的存取。如需 IAM ABAC 的詳細資訊，請參閱《IAM 使用者指南》中的什麼是 ABAC AWS？。

例如，您可以使用 `project-access` 標籤鍵建立三個角色。將第一個角色的鍵值設為 `Dev`，第二個角色的鍵值設為 `Marketing`，並將第三個角色的鍵值設為 `Support`。將具有適當值的標籤指派給資源。然後，您可以使用單一政策，在角色和資源針對 `project-access` 使用相同值標記時允許存取。

資料控管團隊使用 Lake Formation 來定義特定資料湖資源的精細許可。LF 標籤會指派給 Data Catalog 資源（資料庫、資料表和資料欄），並授予委託人。具有符合資源 LF 標籤之 LF 標籤的主體可以存取該資源。Lake Formation 許可是 IAM 許可的次要許可。例如，如果 IAM 許可不允許使用者存取資料湖，Lake Formation 不會將該資料湖中任何資源的存取權授予該使用者，即使委託人和資源具有相符的 LF 標籤。

Lake Formation 標籤型存取控制 (LF-TBAC) 與 IAM ABAC 搭配使用，為您的 Lake Formation 資料和資源提供額外的許可層級。

- Lake Formation TBAC 許可會隨著創新而擴展。管理員不再需要更新現有政策來允許存取新的資源。例如，假設您使用 IAM ABAC 策略搭配 `project-access`標籤，以提供對 Lake Formation 中特定資料庫的存取。使用 LF-TBAC，LF-Tag `Project=SuperApp`會指派給特定資料表或資料欄，而相同的 LF-Tag 會授予該專案的開發人員。透過 IAM，開發人員可以存取資料庫，而 LF-TBAC 許可會授予開發人員進一步存取資料表內特定資料表或資料欄的權限。如果將新資料表新增至專案，Lake Formation 管理員只需要將標籤指派給新資料表，開發人員才能存取資料表。
- Lake Formation TBAC 需要較少的 IAM 政策。由於您使用 IAM 政策授予 Lake Formation 資源和 Lake Formation TBAC 的高階存取權，以管理更精確的資料存取，因此您可以建立較少的 IAM 政策。
- 使用 Lake Formation TBAC，團隊可以快速變更和成長。這是因為新資源會自動根據屬性授與許可。例如，如果新的開發人員加入專案，則將 IAM 角色與使用者建立關聯，然後將必要的 LF 標籤指派給使用者，即可輕鬆授予此開發人員存取權。您不需要變更 IAM 政策來支援新專案或建立新的 LF 標籤。

- 使用 Lake Formation TBAC 可以獲得更精細的許可。IAM 政策會授予最上層資源的存取權，例如 Data Catalog 資料庫或資料表。使用 Lake Formation TBAC，您可以授予存取包含特定資料值的特定資料表或資料欄。

> **Note**
>
> IAM 標籤與 LF 標籤不同。這些標籤不可互換。LF 標籤用於授予 Lake Formation 許可，而 IAM 標籤用於定義 IAM 政策。

管理中繼資料存取控制的 LF-Tag 表達式

LF-Tag 表達式是由一或多個 LF-Tag（鍵/值對）組成的邏輯表達式，用於授予 AWS Glue Data Catalog 資源的許可。LF-Tag 表達式可讓您定義規則，根據資料資源的中繼資料標籤來控管對資料資源的存取。您可以儲存這些表達式，並在多個許可授予中重複使用這些表達式，確保一致性，並使其直接管理標籤本體隨時間的變化。

在指定的 LF-Tag 表達式中，標籤索引鍵是使用 AND 操作組合，而值是使用 OR 操作組合。例如，標籤表達式`content_type:Sales AND location:US`代表與美國銷售資料相關的資源。

您可以在 中建立最多 1000 個 LF-Tag 表達式 AWS 帳戶。這些表達式提供彈性且可擴展的方式，以根據中繼資料標籤管理許可，確保只有授權的使用者或應用程式才能根據定義的標籤規則存取特定資料資源。

LF-Tag 表達式提供下列優點：

- 可重複使用性 – 透過定義和儲存 LF-Tag 表達式，您不再需要在將許可指派給其他資源或主體時手動複寫相同的表達式。
- 一致性 – 在多個許可授予之間重複使用 LF-Tag 表達式可確保授予和管理許可的方式一致。
- 標籤本體管理 – LF-Tag 表達式有助於管理標籤本體隨時間的變更，因為您可以更新已儲存的表達式，而不是修改個別許可授予。

如需標籤型存取控制的詳細資訊，請參閱 Lake Formation 標籤型存取控制。

LF-Tag 表達式建立者

LF-Tag 表達式建立者是有權建立和管理 LF-Tag 表達式的主體。資料湖管理員可以使用 Lake Formation 主控台、CLI、API 或 SDK 新增 LF-Tag 表達式建立者。LF-Tag 表達式建立者具有隱含

Lake Formation 許可，可建立、更新和刪除 LF-Tag 表達式，以及將 LF-Tag 表達式許可授予其他主體。

非資料湖管理員的 LF-Tag 表達式建立者只會針對其建立的表達式接收隱含 `Alter`、`Describe`、`Drop`和 `Grant with LF-Tag expression`許可。

資料湖管理員也可以授予 LF-Tag 表達式建立者可授予的`Create LF-Tag expression`許可。然後，LF-Tag 表達式建立者可以將建立 LF-Tag 表達式的許可授予其他主體。

主題

- 建立 LF-Tag 表達式所需的 IAM 許可
- 新增 LF-Tag 表達式建立者
- 建立 LF-Tag 表達式
- 更新 LF-Tag 表達式
- 刪除 LF-Tag 表達式
- 列出 LF-Tag 表達式

另請參閱

- 管理 LF-Tag 值許可
- 使用 LF-TBAC 方法授予資料湖許可
- Lake Formation 標籤型存取控制

建立 LF-Tag 表達式所需的 IAM 許可

您必須設定許可，以允許 Lake Formation 主體建立 LF-Tag 表達式。將下列陳述式新增至需要成為 LF-Tag 表達式建立者的主體的許可政策。

Note

雖然資料湖管理員具有隱含 Lake Formation 許可來建立、更新和刪除 LF-Tag 和 LF-Tag 表達式、將 LF-Tag 指派給資源，以及將 LF-Tag 和 LF-Tag 表達式許可授予主體，但資料湖管理員也需要下列 IAM 許可。

如需詳細資訊，請參閱Lake Formation 角色和 IAM 許可參考。

```
{
"Sid": "Transformational",
"Effect": "Allow",
    "Action": [
        "lakeformation:AddLFTagsToResource",
        "lakeformation:RemoveLFTagsFromResource",
        "lakeformation:GetResourceLFTags",
        "lakeformation:ListLFTags",
        "lakeformation:CreateLFTag",
        "lakeformation:GetLFTag",
        "lakeformation:UpdateLFTag",
        "lakeformation:DeleteLFTag",
        "lakeformation:SearchTablesByLFTags",
        "lakeformation:SearchDatabasesByLFTags",
        "lakeformation:CreateLFTagExpression",
        "lakeformation:DeleteLFTagExpression",
        "lakeformation:UpdateLFTagExpression",
        "lakeformation:GetLFTagExpression",
        "lakeformation:ListLFTagExpressions",
        "lakeformation:GrantPermissions",
        "lakeformation:RevokePermissions",
        "lakeformation:BatchGrantPermissions",
        "lakeformation:BatchRevokePermissions"
     ]
 }
```

新增 LF-Tag 表達式建立者

LF-Tag 表達式建立者可以建立和儲存可重複使用的 LF-Tag 表達式、更新標籤索引鍵和值、刪除表達式，以及使用 LF-TBAC 方法將 Data Catalog 資源的許可授予委託人。LF-Tag 表達式建立者也可以將這些許可授予委託人。

您可以使用 AWS Lake Formation 主控台、 API 或 AWS Command Line Interface () 來建立 LF-Tag 表達式建立者角色AWS CLI。

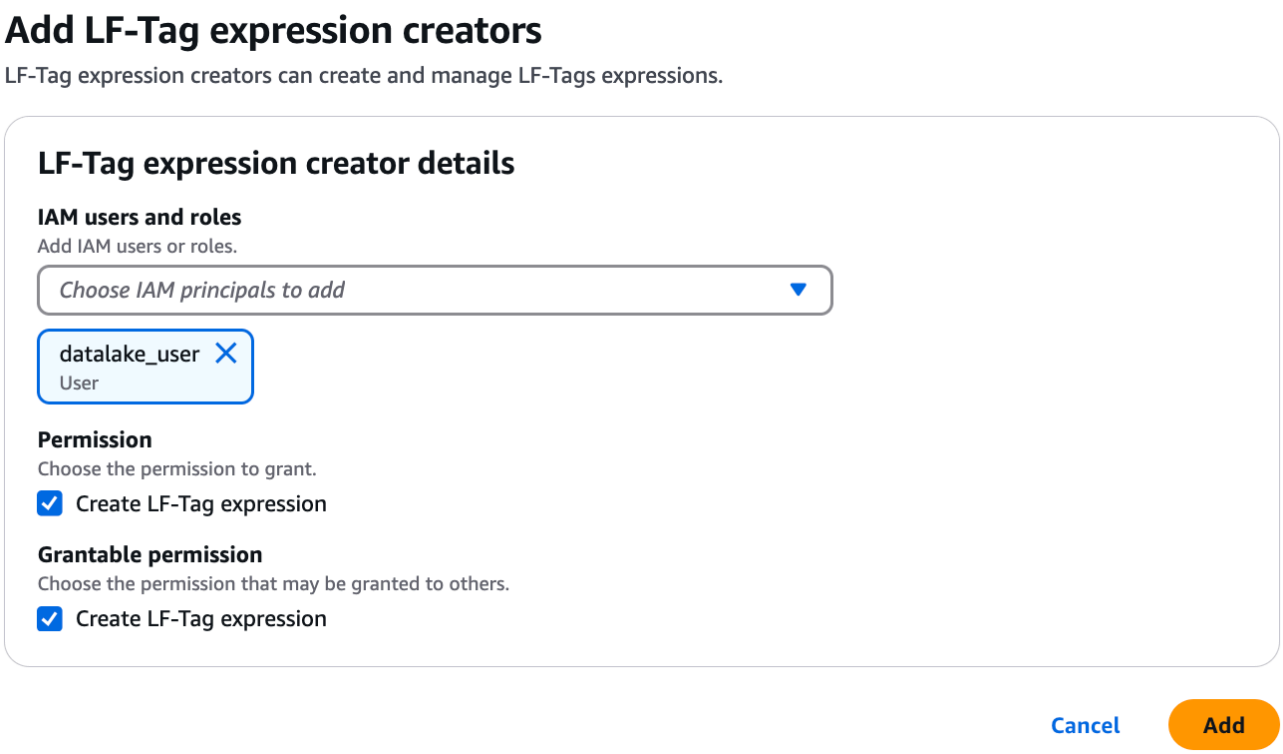
console

新增 LF-Tag 表達式建立者

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

   以資料湖管理員身分登入。

2. 在導覽窗格的許可下，選擇 LF 標籤和許可。
3. 選擇 LF-Tag 表達式索引標籤。
4. 在 LF-Tag 表達式建立者區段中，選擇新增 LF-Tag 表達式建立者。



5. 在新增 LF-Tag 表達式建立者頁面上，選擇具有建立 LF-Tag 表達式所需許可的 IAM 角色或使用者。
6. 選取`Create LF-Tag expression`許可核取方塊。
7. （選用）若要讓選取的委託人將`Create LF-Tag expression`許可授予委託人，請選擇可授予`Create LF-Tag expression`許可。
8. 選擇新增。

AWS CLI

```
aws lakeformation grant-permissions --cli-input-json file://grantCreate
{
    "Principal": {
        "DataLakePrincipalIdentifier": "arn:aws:iam::123456789012:user/tag-manager"
    },
    "Resource": {
        "Catalog": {}
```

```
    },
    "Permissions": [
        "CreateLFTagExpression"
    ],
    "PermissionsWithGrantOption": [
        "CreateLFTagExpression"
    ]
}
```

LF-Tag 表達式建立者角色可以建立、更新或刪除 LF-Tag 表達式。

權限	描述
`Create`	具有此許可的主體可以在資料湖中新增 LF-Tag 表達式。
`Drop`	在 LF-Tag 表達式上具有此許可的委託人可以從資料湖刪除 LF-Tag 表達式。
`Alter`	在 LF-Tag 表達式上具有此許可的主體可以更新 LF-Tag 表達式的表達式內文。
`Describe`	在 LF-Tag 表達式上具有此許可的主體可以檢視 LF-Tag 表達式的內容。
`Grant with LF-Tag expression`	此許可允許收件人在授予資料或中繼資料存取許可時，使用標籤表達式做為資源。隱`Grant with LF-Tag expression` 含授予 `Describe`。
`Super`	對於 LF-Tag 表達式， `Super`許可會授予 `Describe`、`Drop`、 `Alter`和對標籤表達式的許可給其他主體。

這些許可是可授予的。已使用授予選項授予這些許可的委託人，可以將它們授予其他委託人。

建立 LF-Tag 表達式

您需要在 Lake Formation 中定義所有 LF 標籤，並將它們指派給 Data Catalog 資源，才能用來建立運算式。LF-Tag 表達式包含一或多個索引鍵，以及每個索引鍵的一或多個可能值。

在資料湖管理員設定 LF-Tag 表達式建立者角色所需的 IAM 許可和 Lake Formation 許可之後，主體可以建立可重複使用的 LF-Tag 表達式。LF-Tag 表達式建立者會取得隱含許可來更新表達式內文，並刪除 LF-Tag 表達式。

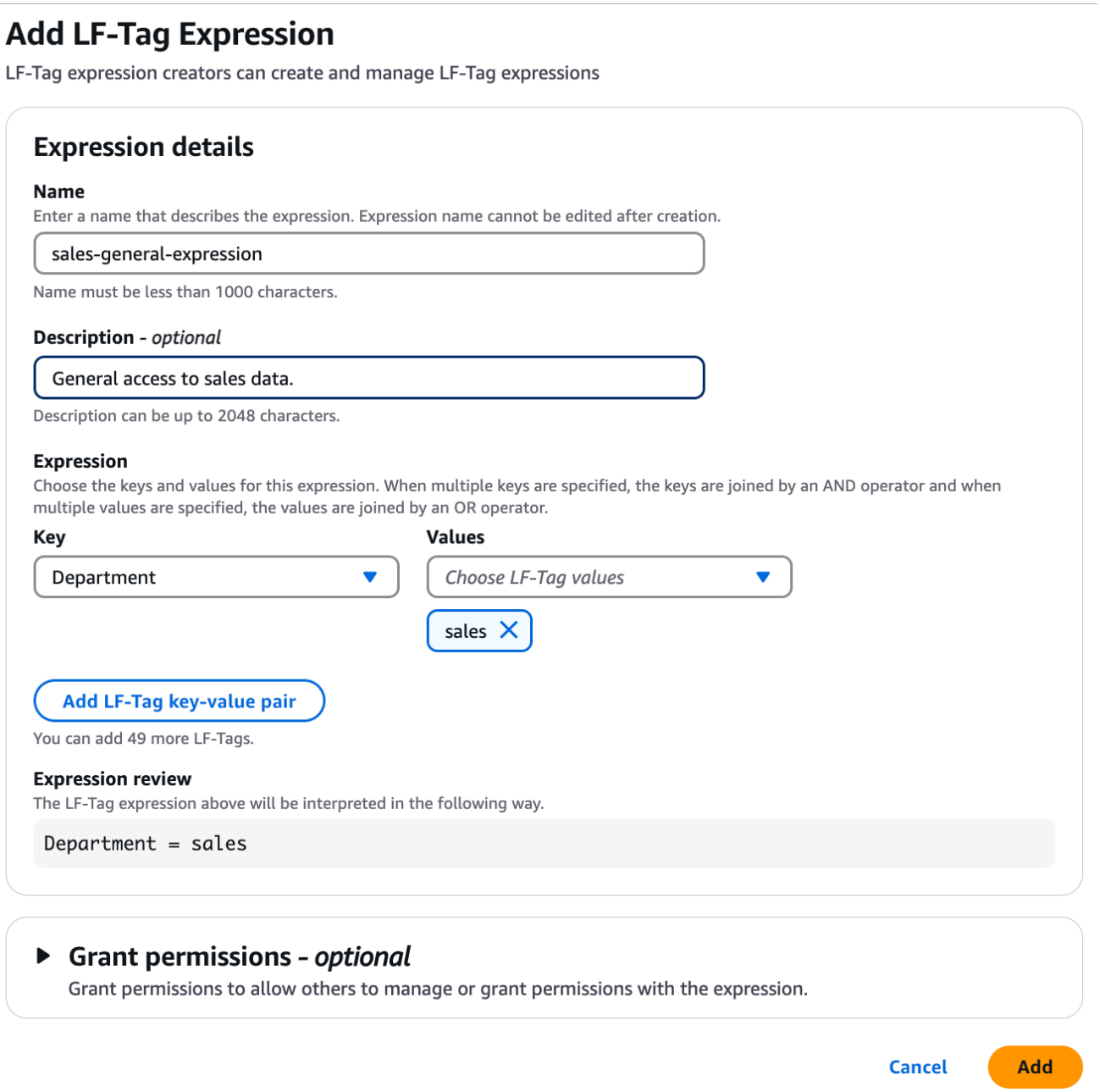
您可以使用 AWS Lake Formation 主控台、 API 或 () 建立 LF-Tag AWS Command Line Interface 表達式AWS CLI。

Console

建立 LF-Tag 表達式

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

   使用 LF-Tag 表達式建立者許可以主體身分登入，或以資料湖管理員身分登入。
2. 在導覽窗格的許可下，選擇 LF 標籤和許可。
3. 選擇 LF-Tag 表達式。新增 LF-Tag 表達式頁面隨即出現。



4. 輸入下列資訊：

   - 名稱 – 輸入表達式的唯一名稱。您無法更新表達式名稱。
   - 描述 – 為表達式提供選用的描述，其中包含表達式的詳細資訊。
   - 表達式 – 透過指定標籤索引鍵及其相關聯的值來建立表達式。每個表達式最多可以新增 50 個索引鍵。您必須擁有表達式內文中所有標籤的 `Grant with LF-Tags` Lake Formation 許可。

每個金鑰必須至少有一個值。若要輸入多個值，請輸入逗號分隔清單，然後按 Enter，或一次輸入一個值，然後選擇在每個值後面新增。每個索引鍵允許的最大值為 1000。

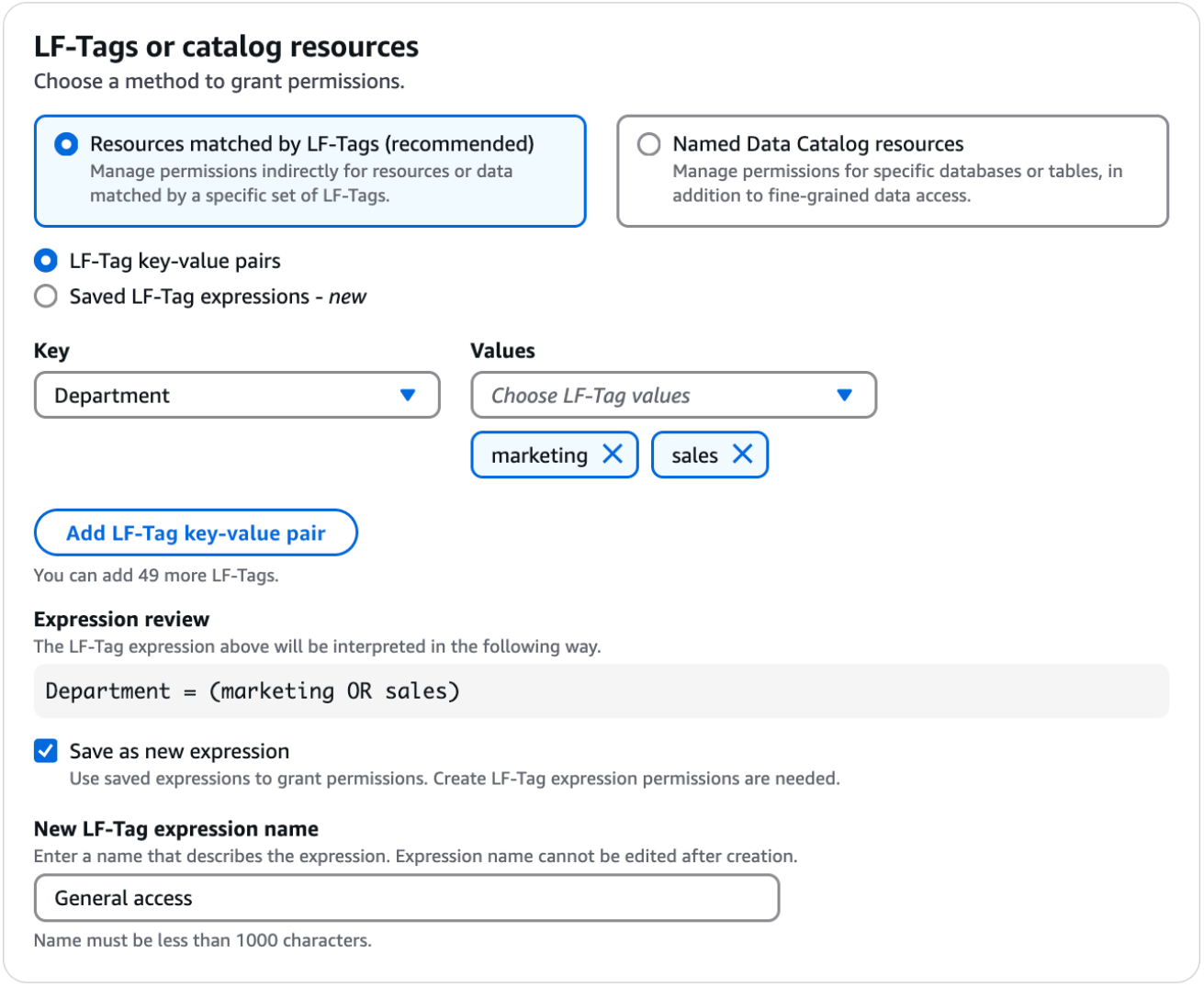
Lake Formation 使用 AND/OR 邏輯，在表達式中結合多個索引鍵和值。在單一（索引鍵：值清單）對中，使用邏輯 OR 運算子來組合值。例如，如果配對是（部門：【銷售、行銷】），則表示如果資源具有銷售或行銷值的部門標籤，則標籤會相符。

當您指定多個金鑰時，金鑰會由 AND 邏輯運算子聯結。因此，如果完整表達式是（部門：【銷售、行銷】) AND（位置：【美國、加拿大】），則會比對具有部門標籤且值為銷售或行銷的資源，以及具有值為美國或加拿大的位置標籤。以下是具有多個索引鍵和值的另一個範例：

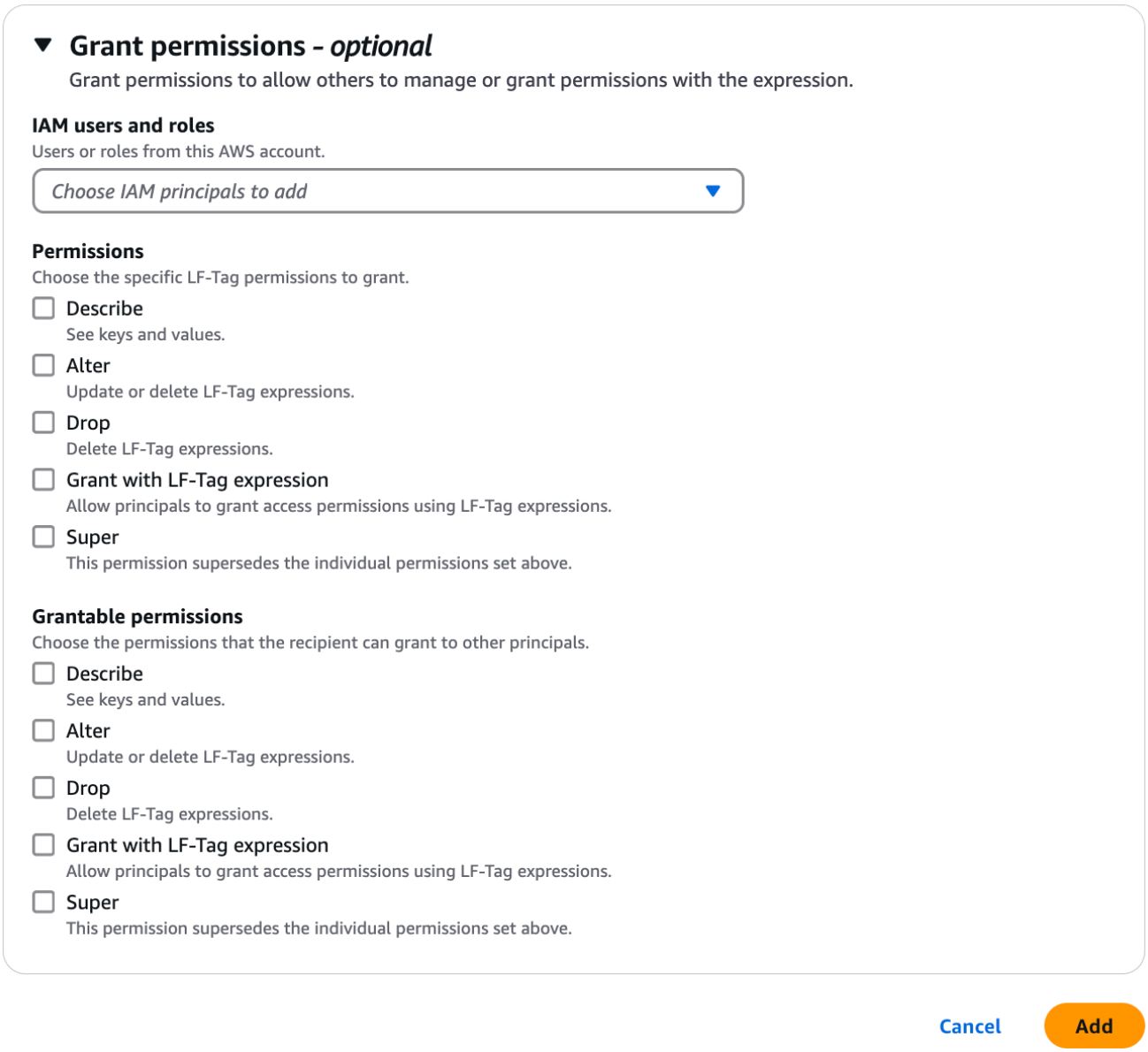
LF-Tag 表達式：(ContentType：【Video， Audio】) AND (Region：【Europe， Asia】) AND (Department：【Engineering， ProductManagement】)。

此表達式會比對具有下列各項的資源：- ContentType 標籤具有值 Video OR Audio AND - 區域標籤具有值 Europe OR Asia AND - 部門標籤具有值 Engineering OR ProductManagement。

您也可以在使用 LF 標籤授予資料湖許可時儲存標籤表達式。選擇索引鍵和值對，然後選擇另存為新表達式選項。輸入描述表達式的名稱。



5. （選用）接著，選擇使用者/角色，以及您想要在帳戶中授予使用者的表達式許可。您也可以選擇可授予的許可，以允許使用者將這些許可授予帳戶中的其他使用者。您無法授予標籤表達式的跨帳戶許可。



6. 選擇新增。

AWS CLI

建立 LF-Tag 表達式

- 輸入`create-lf-tag-expression`命令。

  下列範例會建立 LF-Tag 表達式，其標籤`Department`具有值 `Sales`和 `Marketing`，而標籤`Location`具有值 US。

```
aws lakeformation create-lf-tag-expression \
```

```
-- name "my-tag-expression" \
-- catalog-id "123456789012" \
-- expression '{"Expression":[{"TagKey":"Department","TagValues":
["Sales","Marketing"]},{"TagKey":"Location","TagValues":["US"]}]}'
```

此 CLI 命令會在 中建立新的 LF-Tag 表達式 AWS Glue Data Catalog。表達式可以根據其關聯的標籤，將許可授予 Data Catalog 資源，例如資料庫、資料表、檢視或資料欄。在此範例中，表達式會比對具有值 `Sales`或 之`Department`索引鍵的資源`Marketing`，以及具有值 之索引`Location`鍵的資源US。

作為標籤表達式建立者 ，主體會取得此 LF-Tag 表達式的`Alter`許可，並可以更新或移除表達式。LF-Tag 表達式建立者主體也可以將更新和移除此表達式的`Alter`許可授予另一個主體。

更新 LF-Tag 表達式

只有資料湖管理員、LF-Tag 表達式建立者，以及擁有 LF-Tag 表達式 `Alter` 或 `Super`許可的主體，才能更新 LF-Tag 表達式。除了 `Alter` 許可之外，您也需要新表達式內文上所有基礎索引鍵值的 `lakeformation:UpdateLFTagExpression` IAM 許可和`Grant with LF-Tag`許可，才能更新表達式。

您可以透過更新表達式上授予的描述、表達式內文和許可來更新 LF-Tag 表達式。您無法變更 LF-Tag 表達式的名稱。若要變更名稱，請刪除 LF-Tag 表達式，並使用必要的參數新增。

您可以使用 AWS Lake Formation 主控台、 API 或 AWS Command Line Interface () 來更新 LF-Tag 表達式AWS CLI。

Console

更新 LF-Tag 表達式

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

   以資料湖管理員、LF 標籤建立者或具有 LF 標籤`Alter`許可的主體身分登入。
2. 在導覽窗格中的許可下，選擇 LF 標籤和許可。
3. 選擇 LF-Tag 表達式索引標籤。
4. 在 LF-Tag 表達式區段中，選取 LF-Tag 表達式，然後選擇編輯。
5. 在編輯 LF-Tag 表達式對話方塊中，更新描述，並透過新增或移除索引鍵和值來更新表達式內文。

若要新增多個值，請在值欄位中，從下拉式清單中選擇值。

6. 選擇 Save (儲存)。

AWS CLI

Lake Formation 中的 update-lf-tag-expression 命令可讓您更新現有的 LF-Tag 表達式。

```
aws lakeformation update-lf-tag-expression \
-- name expression_name\
-- description new_description \
-- catalog-id catalog_id \
-- expression '{"Expression": [{"TagKey": "tag_key", "TagValues": ["tag_value1",
 "tag_value2", ...]}]}'
```

以下是所提供命令中的參數所代表的意義：

- name – 您要更新的現有具名標籤表達式的名稱。
- description – 表達式的新描述。

  catalog-id – 具名標籤表達式所在的 Data Catalog ID。
- expression – 您要更新表達式的新標籤表達式字串。

刪除 LF-Tag 表達式

您可以刪除不再使用的 LF-Tag 表達式。如果您已使用 LF-Tag 表達式將許可授予 Data Catalog 資源上的主體，他們將不再擁有許可。

只有資料湖管理員、LF-Tag 表達式建立者，或具有 LF-Tag 表達式`Drop`許可的主體，才能刪除 LF-Tag 表達式。除了 `Drop`許可之外，委託人還需要 `lakeformation:DeleteLFTagExpression` IAM 許可才能刪除 LF-Tag 表達式。

您可以使用 AWS Lake Formation 主控台、 API 或 AWS Command Line Interface () 來刪除 LF-Tag 表達式AWS CLI。

Console

刪除 LF-Tag 表達式（主控台）

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

   以資料湖管理員、LF-Tag 表達式建立者或具有刪除表達式許可的主體身分登入。
2. 在導覽窗格的許可下，選擇 LF 標籤和許可。
3. 選擇 LF-Tag 表達式索引標籤。
4. 在 LF-Tag 表達式區段中，選取 LF-Tag 表達式，然後選擇刪除。
5. 在刪除 LF-Tag 表達式？對話方塊中，若要確認刪除，請在指定欄位中輸入 LF-Tag 表達式名稱，然後選擇刪除。

AWS CLI

刪除 LF 標籤 (AWS CLI)

- 輸入`delete-lf-tag-expression`命令。提供要刪除的表達式名稱和目錄 ID。

  Example

  下列範例`my-tag-expression`會從 ID 為 的資料目錄中刪除名稱為 的 LF-Tag 表達式`123456789012`。如果您使用的是與 AWS CLI 組態相同的帳戶，則 `catalog-id` 參數為選用。刪除 LF-Tag 表達式後，Lake Formation 會清除該表達式的相關聯許可記錄。這包括個別許可記錄和包含已刪除表達式的彙總許可記錄。

  ```
  aws lakeformation delete-lf-tag-expression \
  --name "my-tag-expression" \
  --catalog-id "123456789012"
  ```

列出 LF-Tag 表達式

您可以列出您擁有描述許可的 LF-Tag 表達式。資料湖管理員、LF-Tag 表達式建立者和唯讀管理員可以隱含地查看其帳戶中的所有標籤表達式。

您可以使用 AWS Lake Formation 主控台、 API 或 () 列出 LF-Tag AWS Command Line Interface 表達式AWS CLI。

Console

列出 LF-Tag 表達式（主控台）

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

   以 LF-Tag 表達式建立者身分、資料湖管理員身分，或已授予 LF-Tag 表達式許可且具有 IAM `lakeformation:ListLFTagExpressions` 許可的主體身分登入。
2. 在導覽窗格的許可、LF 標籤和許可下。
3. 選擇 LF-Tag 表達式索引標籤以查看表達式。本節顯示現有 LF-Tag 表達式的相關資訊，包括表達式名稱、包含包含標籤連結的表達式本身，以及建立、編輯或刪除表達式的選項。

AWS CLI

列出 LF 標籤 (AWS CLI)

- 若要使用 列出 LF-Tag 表達式 AWS CLI，您可以使用 list-lf-tag-expressions 命令。請求語法為：

```
aws lakeformation list-lf-tag-expressions \
-- catalog-id "123456789012" \
-- max-items "100" \
-- next-token "next-token"
```

其中：

- `catalog-id` 是您要列出 標籤表達式之 Data Catalog AWS 的帳戶 ID。
- `max-items` 指定要傳回的標籤表達式數目上限。如果未使用此參數，預設值為 100。
- `next-token` 如果在先前的請求中截斷了結果，則 是接續字符。

回應將包含 LF-Tag 表達式的清單，以及下一個字符，如果適用的話。

管理 LF-Tag 值許可

您可以將 LF-Tags 上的 `DropAlter`許可授予委託人，以管理 LF-Tag 值表達式。您也可以將 `Describe`、 `Associate`和 LF 標籤的`Grant with LF-Tag expressions`許可授予委託人，以檢視 LF 標籤並將其指派給 Data Catalog 資源（資料庫、資料表和資料欄）。當 LF 標籤指派給 Data

Catalog 資源時，您可以使用 Lake Formation 標籤型存取控制 (LF-TBAC) 方法來保護這些資源。如需詳細資訊，請參閱Lake Formation 標籤型存取控制。

您可以使用授予選項授予這些許可，以便其他委託人可以授予這些許可。`Grant with LF-Tag expressions`、 `Describe`和 `Associate`許可會在 中說明新增 LF 標籤建立者。

您可以將 LF 標籤的 `Describe`和 `Associate`許可授予外部 AWS 帳戶。然後，該帳戶中的資料湖管理員可以將這些許可授予帳戶中的其他主體。外部帳戶中資料湖管理員授予`Associate`許可的主體接著可以將 LF 標籤指派給您與其帳戶共用的 Data Catalog 資源。

授予外部帳戶時，您必須包含授予選項。

您可以使用 Lake Formation 主控台、 API 或 AWS Command Line Interface () 授予 LF 標籤的許可 AWS CLI。

主題

- 使用主控台列出 LF-Tag 許可
- 使用主控台授予 LF-Tag 許可
- 使用 管理 LF-Tag 許可 AWS CLI

如需更多資訊，請參閱管理中繼資料存取控制的 LF 標籤及Lake Formation 標籤型存取控制。

使用主控台列出 LF-Tag 許可

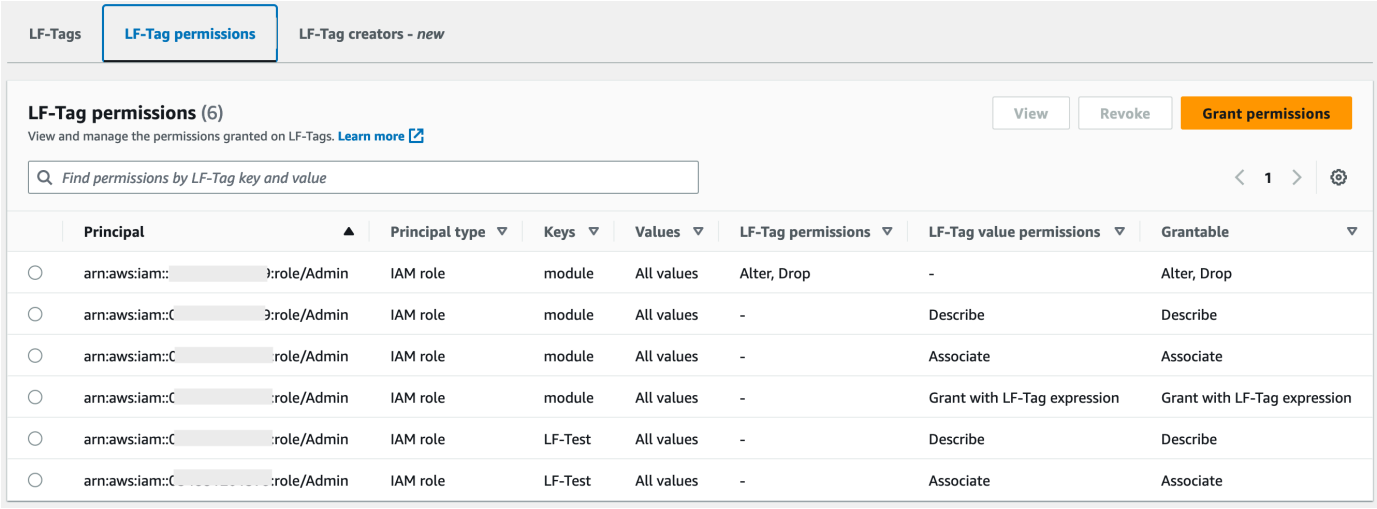
您可以使用 Lake Formation 主控台來檢視 LF-Tags 上授予的許可。您必須是 LF-Tag 建立者、資料湖管理員，或具有 `Describe`或 LF-Tag `Associate`許可才能查看。

列出 LF-Tag 許可 （主控台）

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

   以 LF-Tag 建立者、資料湖管理員或已授予 LF-Tags 上 `Drop`、`Associate`、 `Alter`或 `Describe`許可的使用者身分登入。
2. 在導覽窗格的許可下，選擇 LF 標籤和許可，然後選擇 LF 標籤許可區段。

   LF-Tag 許可區段顯示包含主體、標籤索引鍵、值和許可的資料表。



使用主控台授予 LF-Tag 許可

下列步驟說明如何使用 Lake Formation 主控台上的授予 LF-Tag 許可頁面來授予 LF-Tag 許可。頁面分為以下部分：

- 許可類型 – 要授予的許可類型。
- 委託人 – 要授予許可的 IAM 使用者或角色，或 SAML 使用者或角色。
- LF-Tag 鍵/值對許可 – 要授予許可的 LF-Tag 鍵/值對。
- LF 標籤許可 – 要授予許可的 LF 標籤。
- LF-Tag 表達式許可 – 要授予許可的 LF-Tag。
- 許可 – 要授予的許可。

開啟授予 LF-Tag 許可頁面

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

   已使用 `Grant`選項，以 LF-Tag 建立者、資料湖管理員或使用者 LF-Tag 許可或 LF-Tag 鍵/值對許可的身分登入。
2. 在導覽窗格中，選擇 LF 標籤和許可，選擇 LF 標籤許可區段。
3. 選擇 授予許可。

指定許可類型

在許可類型區段中，選擇許可類型。

LF-Tag 許可

選擇 LF-Tag 許可，以允許主體更新 LF-Tag 值或刪除 LF-Tag。

LF 標籤鍵/值對許可

選擇 LF-Tag 鍵/值對許可，以允許主體將 LF-Tags 指派給 Data Catalog 資源、檢視 LF-Tags 和值，並將基於 LF-Tags 的許可授予主體。

下列各節中可用的選項取決於許可類型。

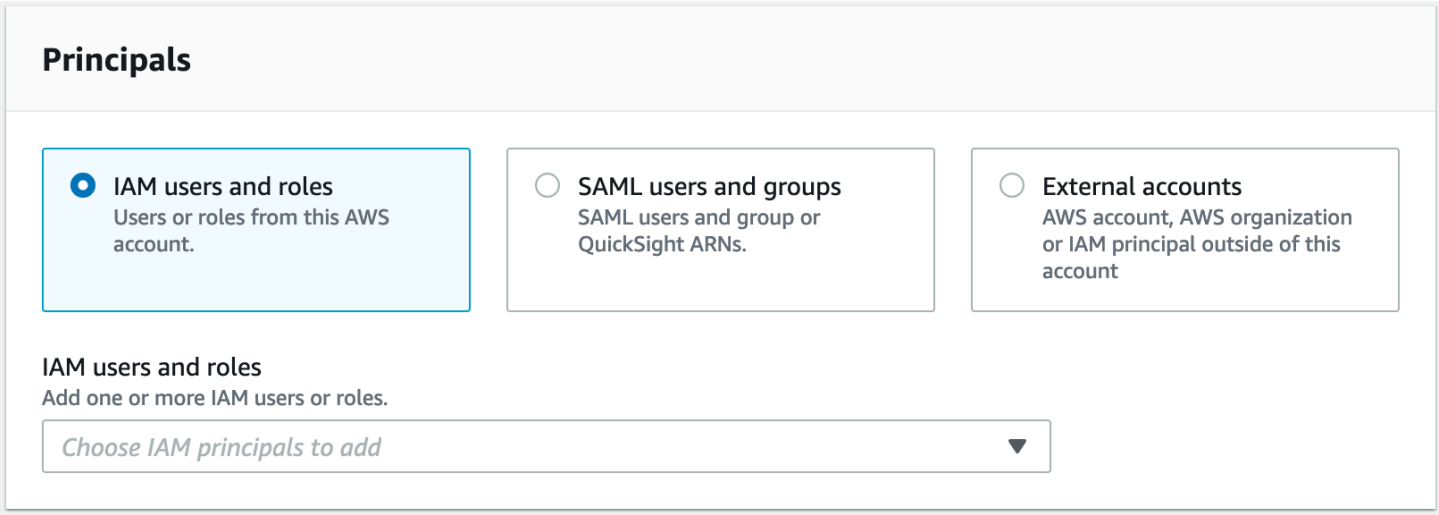
LF-Tag 表達式許可

選擇 LF-Tag 表達式許可，以允許主體更新表達式或刪除表達式。

指定委託人

> **Note**
>
> 您無法將 LF-Tag 許可 (`Alter` 和 `Drop`) 授予另一個帳戶中的外部帳戶或主體。

在委託人區段中，選擇委託人類型，並指定要授予許可的委託人。



IAM 使用者和角色

從 IAM 使用者和角色清單中選擇一或多個使用者或角色。

SAML 使用者和群組

對於 SAML 和 QuickSight 使用者和群組，輸入透過 SAML 聯合的使用者或群組的一或多個 Amazon Resource Name (ARNs)，或 QuickSight 使用者或群組ARNs。在每個 ARN 之後按 Enter。

如需如何建構 ARNs 的資訊，請參閱 Lake Formation 授予和撤銷 AWS CLI 命令。

> **Note**
>
> Lake Formation 與 QuickSight 整合僅支援 QuickSight Enterprise Edition。

外部帳戶

針對AWS 帳戶，輸入一或多個有效的 AWS 帳戶 IDs。在每個 ID 之後按 Enter。

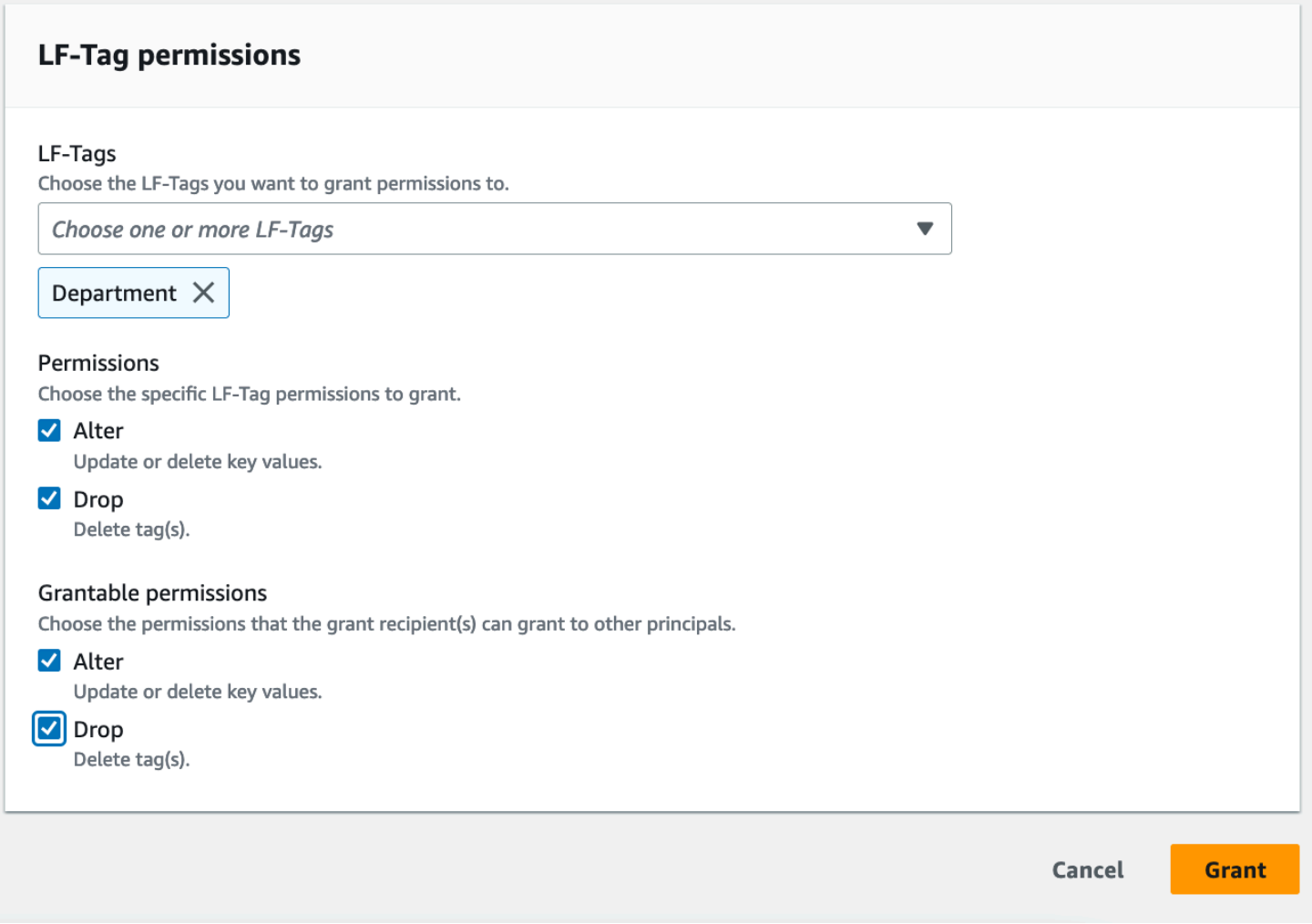
組織 ID 包含「o-」，後面接著 10 到 32 個小寫字母或數字。

組織單位 ID 以「ou-」開頭，後面接著 4 到 32 個小寫字母或數字（包含 OU 的根 ID)。此字串後面接著第二個「-」破折號和 8 到 32 個額外的小寫字母或數字。

針對 IAM 主體，輸入 IAM 使用者或角色的 ARN。

指定 LF 標籤

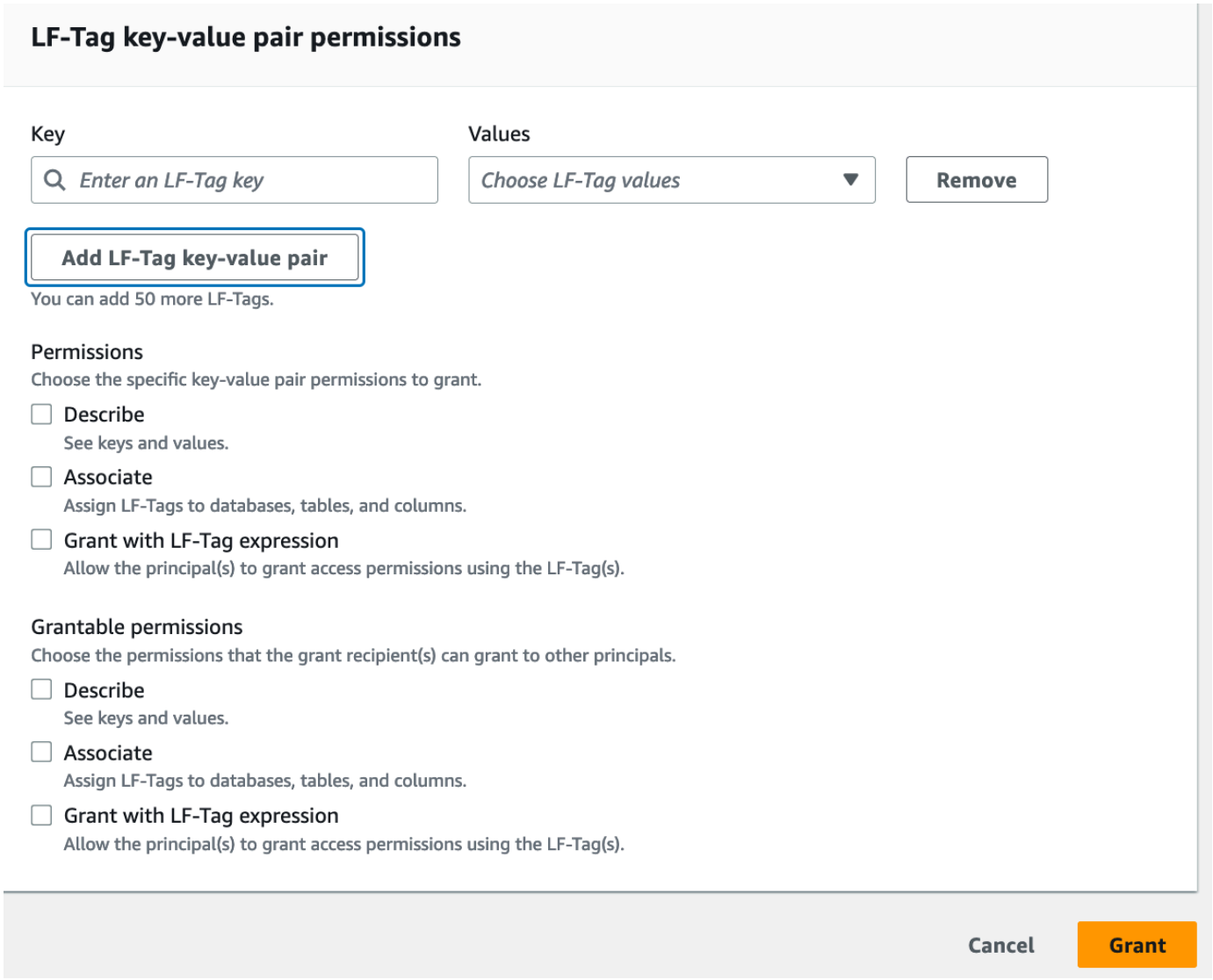
若要授予 LF 標籤的許可，請在 LF 標籤許可區段中指定要授予許可的 LF 標籤。



- 使用下拉式清單選擇一或多個 LF 標籤。

指定 LF 標籤鍵/值對

1. 若要授予 LF-Tag 鍵/值對的許可，（您必須先選擇 LF-Tag 鍵/值對許可做為許可類型) 選擇新增 LF-Tag 鍵/值對，以顯示用於指定 LF-Tag 鍵和值的第一列欄位。



2. 將游標放在索引鍵欄位中，選擇性地開始輸入以縮小選取範圍清單，然後選取 LF-Tag 索引鍵。
3. 在值清單中，選取一或多個值，然後按下 Tab 或按一下或點選欄位外部以儲存選取的值。

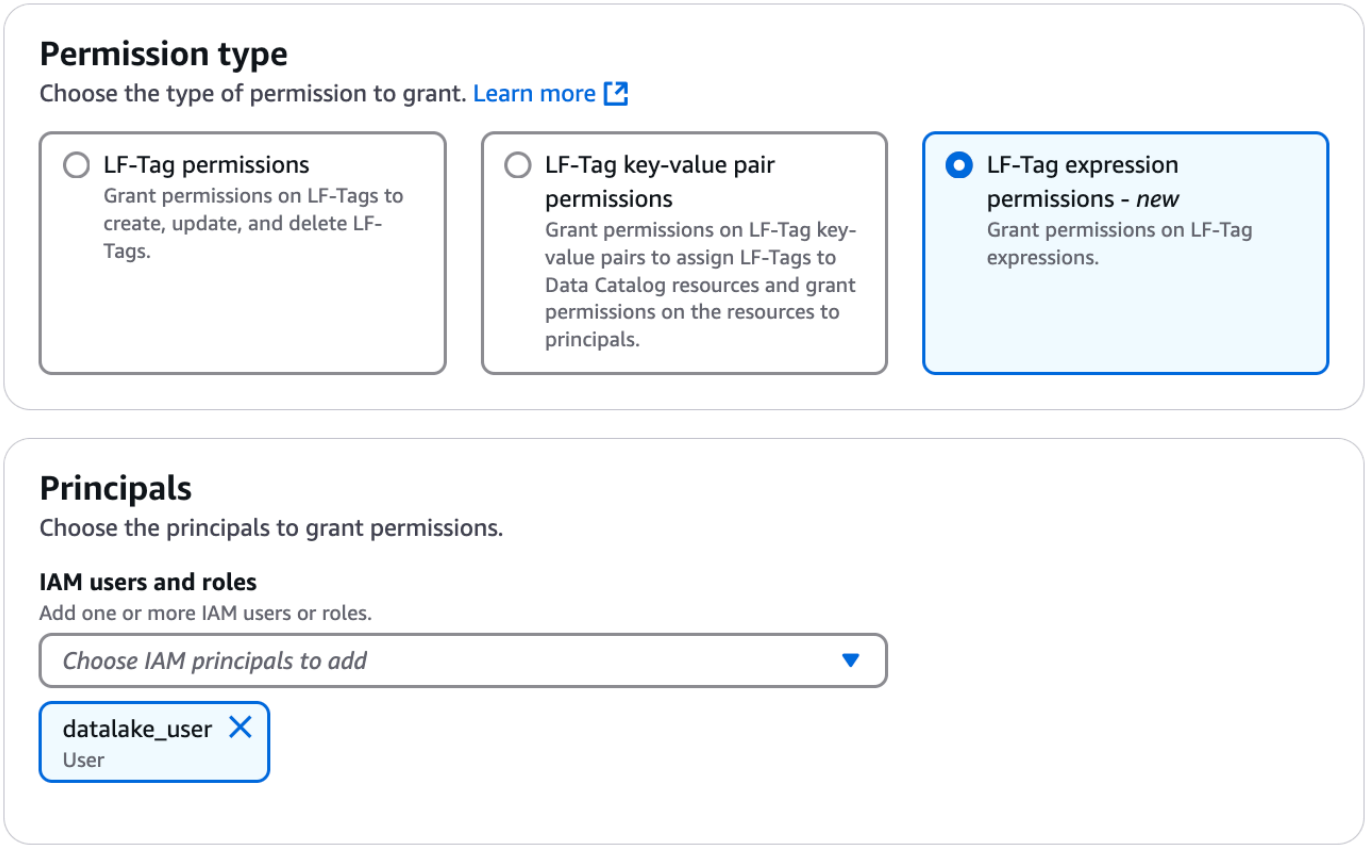
> **Note**
>
> 如果值清單中的其中一個資料列具有焦點，請按 Enter 選取或清除核取方塊。

選取的值會顯示為值清單下方的圖磚。選擇 ✖ 以移除值。選擇移除以移除整個 LF 標籤。

4. 若要新增另一個 LF-Tag，請再次選擇新增 LF-Tag，然後重複前兩個步驟。

指定 LF-Tag 表達式

1. 若要授予 LF-Tag 表達式的許可，您需要先選擇 LF-Tag 表達式許可作為許可類型)。



2. 選擇 LF-Tag 表達式。
3. 選取的表達式會在 LF-Tag 表達式清單下方顯示為圖磚。選擇 ✖ 以移除表達式。
4. 若要新增另一個 LF-Tag 表達式，請選擇另一個表達式。

指定許可

本節會根據您在上一個步驟中選擇的許可類型，顯示 LF-Tag 許可或 LF-Tag 值許可。

根據您選擇授予的許可類型，選取 LF-Tag 許可或 LF-Tag 鍵/值對許可，以及可授予的許可。

1. 在 LF-Tag 許可下，選取要授予的許可。

   授予 Drop and Alter 隱含授予 描述。

   您需要授予所有標籤值的更改和捨棄許可。
2. 在 LT-Tag 鍵值許可下，選取要授予的許可。

授予關聯隱含授予描述。選擇使用 LF-Tag 表達式授予，以允許授予收件人使用 LF-TBAC 方法授予或撤銷 Data Catalog 資源的存取許可。

3. 在 LF-Tag 表達式許可下，選取要授予的許可。

   授予 Drop and Alter 隱含授予 描述。

   授予超級許可， 授予所有可用的許可。

4. （選用）在可授予許可下，選取授予收件人可以授予其 AWS 帳戶中其他主體的許可。
5. 選擇 Grant (授予)。

使用 管理 LF-Tag 許可 AWS CLI

您可以使用 () 授予、撤銷和列出 LF 標籤 AWS Command Line Interface 的許可AWS CLI。

列出 LF-Tag 許可 (AWS CLI)

- 輸入`list-permissions`命令。您必須是 LF-Tag 建立者、資料湖管理員，或具有 `Drop`LF-Tag 的 `Alter`、`Describe`、`Associate`、、 `Grant with LF-Tag permissions`許可才能查看它。

  下列命令會請求您擁有許可的所有 LF 標籤。

  ```
  aws lakeformation list-permissions --resource-type LF_TAG
  ```

  以下是資料湖管理員的範例輸出，該管理員會看到授予所有主體的所有 LF 標籤。非管理使用者只會看到授予給他們的 LF 標籤。從外部帳戶授予的 LF-Tag 許可會顯示在單獨的結果頁面上。若要查看它們，請重複 命令，並使用從上一個命令執行傳回的字符來提供`--next-token`引數。

  ```
  {
      "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"
}
```

您可以列出特定 LF-Tag 金鑰的所有授予。下列命令會傳回 LF 標籤 上授予的所有許可module。

```
aws lakeformation list-permissions --resource-type LF_TAG --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

您也可以列出針對特定 LF 標籤授予特定委託人的 LF 標籤值。提供--principal引數時，您必須提供引--resource數。因此，命令只能針對特定 LF-Tag 金鑰有效地請求授予特定委託人的值。下列命令顯示如何對委託人 datalake_user1和 LF-Tag 金鑰 執行此操作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":["*"]}}'
```

下列為範例輸出。

```
{
    "PrincipalResourcePermissions": [
        {
            "Principal": {
                "DataLakePrincipalIdentifier": "arn:aws:iam::111122223333:user/
datalake_user1"
            },
            "Resource": {
                "LFTag": {
                    "CatalogId": "111122223333",
                    "TagKey": "module",
                    "TagValues": [
                        "Orders",
                        "Sales"
                    ]
                }
            },
            "Permissions": [
                "ASSOCIATE"
            ],
            "PermissionsWithGrantOption": []
        }
    ]
}
```

授予 LF-Tags 的許可 (AWS CLI)

1. 輸入與以下相似的命令。此範例datalake_user1使用金鑰 授予使用者 LF-Tag 上的Associate許可module。它授予檢視和指派該索引鍵所有值的許可，如星號 (*) 所示。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --permissions "ASSOCIATE" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

授予Associate許可隱含授予Describe許可。

下一個範例使用 金鑰 Associate ，搭配 授予選項module，在 LF-Tag 上授予外部 AWS 帳戶 1234-5678-9012。它授予僅檢視和指派值 sales和 的許可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. 隱含授予GrantWithLFTagExpression許可會隱含授予Describe許可。

   下一個範例使用 金鑰 GrantWithLFTagExpression將 LF 標籤上的使用者授予 module，並使用 授予選項。它只授予使用值 和 來檢視sales和授予 Data Catalog 資源許可的許可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. 下一個範例使用 金鑰 將Drop許可授予 LF-Tag 上的使用者module，並使用 授予選項。它授予刪除 LF 標籤的許可。若要刪除 LF 標籤，您需要該金鑰所有值的許可。

   ```
   aws lakeformation grant-permissions --principal
    DataLakePrincipalIdentifier=111122223333 --permissions "DROP"
    --permissions-with-grant-option "DROP" --resource '{ "LFTag":
    {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
   ```

4. 下一個範例使用 金鑰 將 LF-Tag 上的Alter許可授予使用者module，並使用 授予選項。它授予刪除 LF 標籤的許可。若要更新 LF 標籤，您需要該金鑰所有值的許可。

   ```
   aws lakeformation grant-permissions --principal
    DataLakePrincipalIdentifier=111122223333 --permissions "ALTER"
    --permissions-with-grant-option "ALTER" --resource '{ "LFTag":
    {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
   ```

撤銷 LF 標籤的許可 (AWS CLI)

- 輸入與以下相似的命令。此範例會撤銷 LF 標籤的Associate許可，其中包含module來自使用者的 金鑰datalake_user1。

```
aws lakeformation revoke-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/
datalake_user1 --permissions "ASSOCIATE" --resource '{ "LFTag":
 {"CatalogId":"111122223333","TagKey":"module","TagValues":["*"]}}'
```

使用 LF-TBAC 方法授予資料湖許可

您可以授予委託人 LF 標籤的 DESCRIBE和 ASSOCIATE Lake Formation 許可，讓他們可以檢視 LF 標籤，並將其指派給 Data Catalog 資源 （資料庫、資料表、檢視和資料欄）。當 LF 標籤指派給 Data Catalog 資源時，您可以使用 Lake Formation 標籤型存取控制 (LF-TBAC) 方法來保護這些資源。如需詳細資訊，請參閱Lake Formation 標籤型存取控制。

一開始，只有資料湖管理員可以授予這些許可。如果資料湖管理員使用授予選項授予這些許可，其他主體可以授予這些許可。DESCRIBE 和 ASSOCIATE許可會在 中說明Lake Formation 標籤型存取控制最佳實務和考量事項。

您可以將 LF 標籤的 DESCRIBE和 ASSOCIATE許可授予外部 AWS 帳戶。然後，該帳戶中的資料湖管理員可以將這些許可授予帳戶中的其他主體。外部帳戶中資料湖管理員授予ASSOCIATE許可的主體，接著可以將 LF 標籤指派給您與其帳戶共用的 Data Catalog 資源。

授予外部帳戶時，您必須包含授予選項。

您可以使用 AWS Lake Formation 主控台、 API 或 AWS Command Line Interface () 授予 LF 標籤的許可AWS CLI。

主題

- 授予 Data Catalog 許可

> **另請參閱**
>
> - 管理 LF-Tag 值許可
> - 管理中繼資料存取控制的 LF 標籤

- Lake Formation 標籤型存取控制

授予 Data Catalog 許可

使用 Lake Formation 主控台或使用 Lake Formation 標籤型存取控制 AWS CLI (LF-TBAC) 方法來授予資料目錄資料庫、資料表、檢視和資料欄的 Lake Formation 許可。

Console

下列步驟說明如何使用 Lake Formation 標籤型存取控制 (LF-TBAC) 方法和 Lake Formation 主控台上的授予資料湖許可頁面來授予許可。頁面分為以下部分：

- 委託人 – AWS 帳戶 要授予許可的使用者、角色和 。
- LF 標籤或目錄資源 – 要授予許可的資料庫、資料表或資源連結。
- 許可 – 要授予的 Lake Formation 許可。

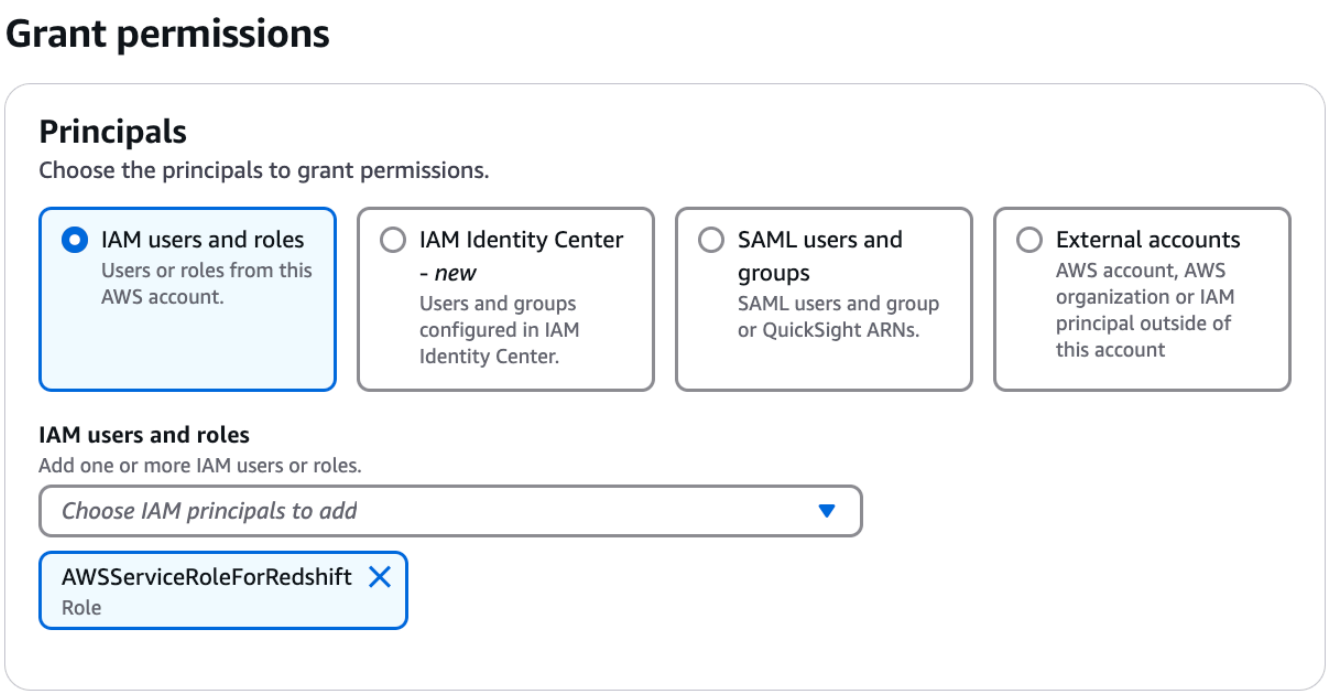
1. 開啟授予資料湖許可頁面。

   開啟位於 https：//https://console.aws.amazon.com/lakeformation/ 的 AWS Lake Formation 主控台，並以資料湖管理員或已透過 LF-TBAC 授予資料目錄資源 Lake Formation 許可的使用者身分登入。

   在導覽窗格中的許可下，選擇資料湖許可。然後選擇授予。

2. 指定委託人。

   在委託人區段中，選擇委託人類型，然後指定要授予許可的委託人。



IAM 使用者和角色

從 IAM 使用者和角色清單中選擇一或多個使用者或角色。

IAM Identity Center

從使用者和群組清單中選擇一或多個使用者或 。

SAML 使用者和群組

對於 SAML 和 QuickSight 使用者和群組，輸入透過 SAML 聯合的使用者或群組的一或多個 Amazon Resource Name (ARNs)，或 QuickSight 使用者或群組ARNs。在每個 ARN 之後按 Enter。

如需如何建構 ARNs 的資訊，請參閱 Lake Formation 授予和撤銷 AWS CLI 命令。

> **Note**
>
> 僅 QuickSight Enterprise Edition 支援 Lake Formation 與 QuickSight 整合。

外部帳戶

針對 AWS 帳戶、 AWS organization 或 IAM 主體，輸入 IAM 使用者或角色的一或多個有效 AWS 帳戶 IDs、組織 IDs、組織單位 IDs 或 ARN。在每個 ID 之後按 Enter。

組織 ID 包含「o-」，後面接著 10 到 32 個小寫字母或數字。

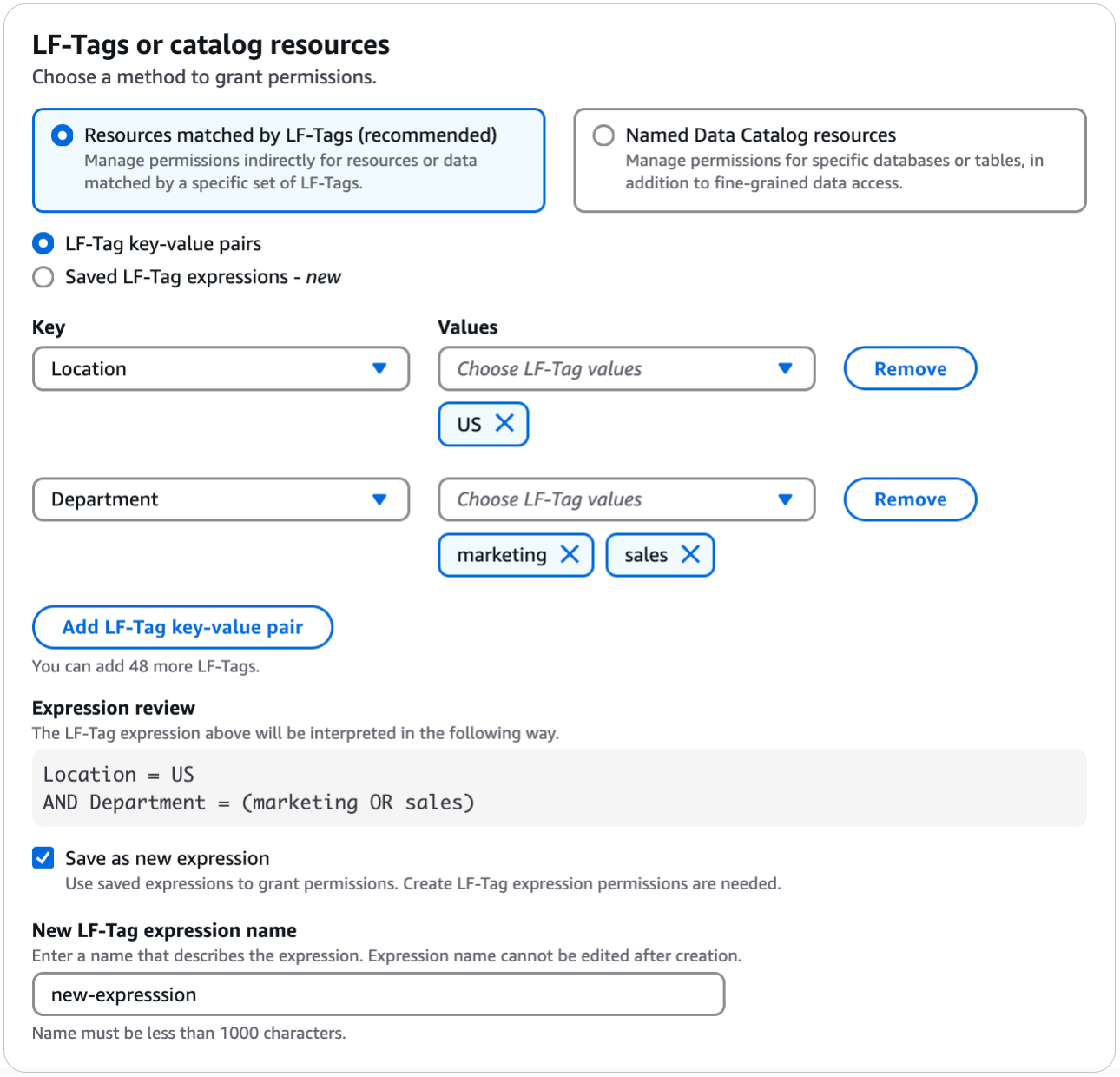
組織單位 ID 以「ou-」開頭，後面接著 4 到 32 個小寫字母或數字（包含 OU 的根 ID)。此字串後面接著第二個「-」破折號和 8 到 32 個額外的小寫字母或數字。

3. 指定 LF 標籤。

確定已選擇符合 LF-Tags 的資源選項。選擇 LF-Tag 鍵/值對或已儲存的 LF-Tag 表達式。

1. 如果您選擇 LF-Tag 鍵/值對選項，請選擇鍵和值。

如果您選擇多個值，則會使用 OR運算子建立 LF-Tag 表達式。這表示如果任何 LF-Tag 值符合指派給 Data Catalog 資源的 LF-Tag，則會授予您資源的許可。



2. （選用）再次選擇新增 LF 標籤鍵/值對，以指定另一個 LF 標籤。

   如果您指定多個 LF-Tag，則會使用 AND運算子建立 LF-Tag 表達式。只有在為 LF-Tag 表達式中的每個 LF-Tag 指派相符的 LF-Tag 時，才會對 Data Catalog 資源授予主體許可。

3. 選擇另存為新的表達式選項以重複使用該表達式。

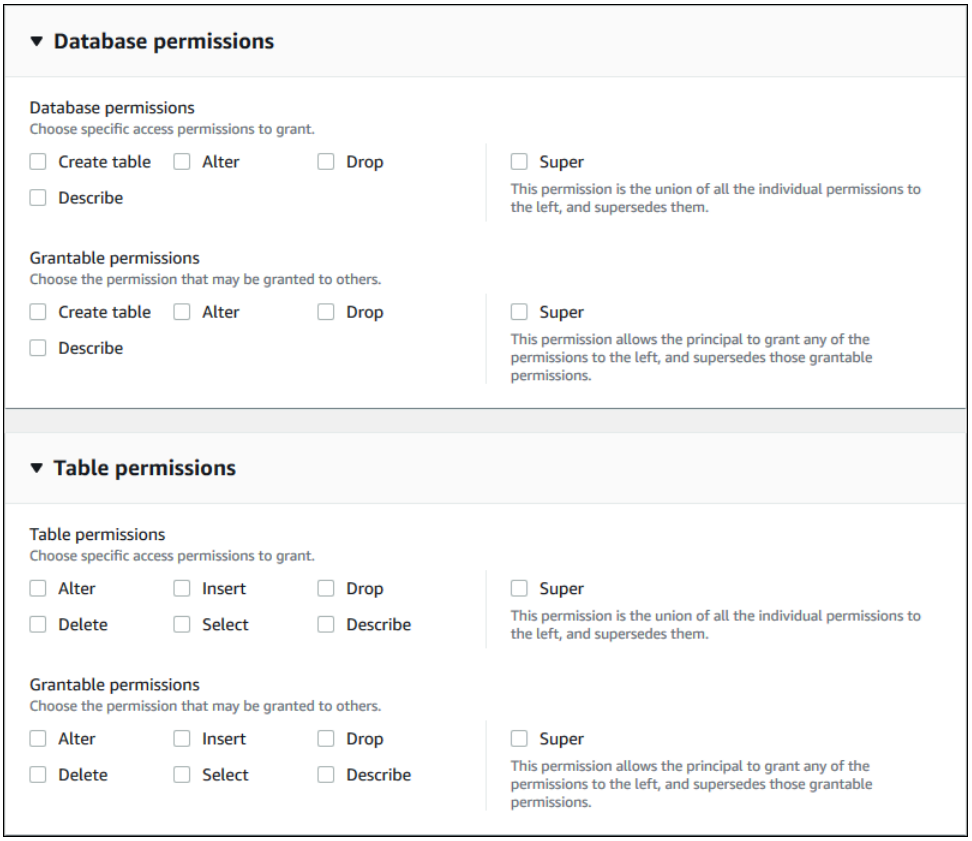
   您需要`Create LF-Tag expression`儲存表達式。

   如需 LF-Tag 表達式的詳細資訊，請參閱 管理中繼資料存取控制的 LF-Tag 表達式。

4. 指定許可。

   指定您希望授予主體相符 Data Catalog 資源的許可。相符資源是指派的 LF 標籤資源，這些資源符合授予委託人的其中一個 LF 標籤表達式。

   您可以指定在相符資料庫、相符資料表和相符檢視上授予 的許可。



   在資料庫許可下，選取要授予相符資料庫上主體的資料庫許可。

   在資料表許可下，選取資料表或檢視許可，以授予相符資料表和檢視的委託人。

   您也可以從資料表`Drop`許可中選擇 `Describe`、 `Select`和 許可，以套用至檢視。

5. 選擇 Grant (授予)。

AWS CLI

您可以使用 AWS Command Line Interface (AWS CLI) 和 Lake Formation 標籤型存取控制 (LF-TBAC) 方法來授予資料目錄資料庫、資料表和資料欄的 Lake Formation 許可。

使用 AWS CLI 和 LF-TBAC 方法授予資料湖許可

- 使用 grant-permissions 命令。

  Example

  下列範例會將 LF-Tag 表達式 "module=*" (LF-Tag 金鑰 的所有值module) 授予使用者 datalake_user1。該使用者將擁有所有相符資料庫的 CREATE_TABLE 許可，這些資料庫已使用索引鍵 指派 LF-Tagmodule，且具有任何值。

```
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

  下一個範例會將 LF-Tag 表達式 "(level=director) AND (region=west OR region=south)" 授予使用者 datalake_user1。該使用者將擁有 SELECT、 和 DROP許可ALTER，並在相符資料表上具有授予選項 - 已同時指派 level=director和 (region=west 或) 的資料表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

  下一個範例會將 LF-Tag 表達式 "module=orders" 授予 AWS 帳戶 1234-5678-9012。然後，該帳戶中的資料湖管理員可以將「module=orders」表達式授予其帳戶中的主體。然後，這些委託人將具有帳戶 1111-2222-3333 所擁有並與帳戶 1234-5678-9012 共用的相符資料庫的CREATE_TABLE許可，方法是使用具名資源方法或 LF-TBAC 方法。

```
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"]}]}}'
```

屬性型存取控制

在 AWS Lake Formation 中，您可以使用與角色和使用者等 IAM 實體相關聯的 IAM 標籤和工作階段標籤屬性，授予目錄、資料庫、資料表和資料篩選條件等 AWS Glue Data Catalog 物件的存取權。

如需使用工作階段標籤的詳細資訊，請參閱《 AWS CLI 使用者指南》中的擔任角色。

屬性型存取控制 (ABAC) 是一種授權策略，可根據屬性定義許可。 會 AWS 呼叫這些屬性標籤。您可以使用 ABAC 將存取權授予相同帳戶或 Data Catalog 資源 上另一個帳戶中的主體。任何具有相符 IAM 標籤或工作階段標籤索引鍵和值的 IAM 主體都可以存取資源。您必須擁有資源的可授予許可，才能進行這些授予。

ABAC 可讓您同時將存取權授予多個使用者。當新使用者加入組織時，其對資料的存取可以根據其屬性自動決定，例如其任務職能或部門，而無需管理員手動指派特定角色或許可。透過使用屬性而非角色，ABAC 提供了更簡化且可維護的方式，來管理各種系統和環境的資料存取，最終增強了資料管理和合規性。

如需定義屬性的詳細資訊，請參閱使用 ABAC 授權根據屬性定義許可。

如需限制、考量和支援 AWS 區域的資訊，請參閱 屬性型存取控制考量事項、限制和支援的區域。

主題

- 使用 屬性授予許可的先決條件
- 使用屬性型存取控制授予許可

使用 屬性授予許可的先決條件

若要使用屬性型存取控制 (ABAC) 授予許可，您必須完成下列先決條件：

- 更新 Data Catalog 設定，以啟用 Data Catalog 物件的 Lake Formation 許可。如需詳細資訊，請參閱變更預設許可模型或使用混合存取模式一節。
- 將跨帳戶版本設定設為兩個或更高版本。
- 將屬性連接至需要存取的 IAM 實體。

- 只有具有必要許可的資料湖管理員或 IAM 使用者才能授予 Data Catalog 物件的存取權。如需必要許可的詳細資訊，請參閱 IAM 許可。

使用屬性型存取控制授予許可

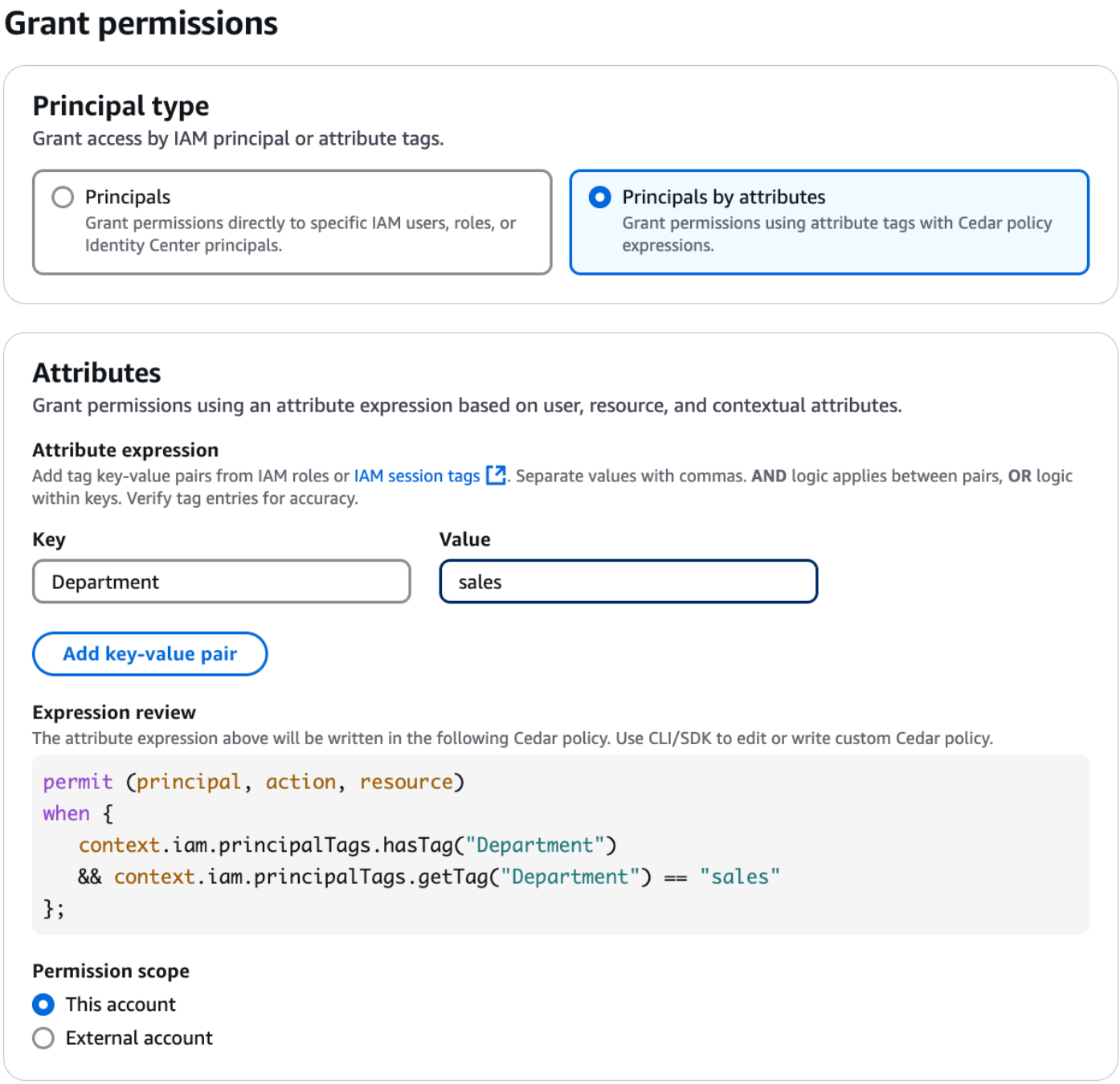
本主題說明您需要遵循的步驟，以授予 Data Catalog 資源的屬性型存取許可。您可以使用 Lake Formation 主控台或 AWS 命令列界面 (AWS CLI)。

使用 ABAC 授予許可 (AWS Management Console)

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/：//。
2. 執行以下任意一項：
   - 在導覽窗格中的許可下，選擇資料湖許可。然後選擇授予。
   - 在導覽窗格中，選擇資料目錄下的目錄。 然後，從許可下的動作選單中選擇目錄物件 （目錄、資料庫、資料表和資料篩選條件），然後選擇授予。
3. 在授予許可頁面上，依屬性選擇主體。
4. 指定屬性索引鍵和值 (s)。如果您選擇多個值，則會使用 OR運算子建立屬性表達式。這表示如果指派給 IAM 角色或使用者配對的任何屬性標籤值，該角色/使用者會取得資源的存取許可。

   如果您指定多個屬性標籤，則會使用 AND運算子建立屬性表達式。只有在屬性表達式中每個屬性標籤的 IAM 角色/使用者獲指派相符標籤時，才會授予 Data Catalog 資源的主體許可。

   檢閱主控台中顯示的產生的 Cedar 政策表達式。



5. 選擇許可範圍。如果承授者屬於外部帳戶，請選擇外部帳戶，然後輸入 AWS 帳戶 ID。
6. 接著，選擇 Data Catalog 帳戶或在外部帳戶中。您必須對資源擁有對應的可授予許可，才能成功完成許可授予。
7. 指定您要允許具有相符屬性之主體（使用者或角色）執行的動作。存取權會授予已指派標籤的 IAM 實體，以及至少符合其中一個指定屬性表達式的值。在 主控台中檢閱 Cedar 政策表達式。如需 Cedar 的詳細資訊，請參閱什麼是 Cedar？| Cedar 政策語言參考GuideLink。

8. 接下來，選擇 Data Catalog 資源以授予存取權。您可以定義各種 Data Catalog 資源的這些許可，包括目錄、資料庫、資料表和資料篩選條件。
9. 選擇 Grant (授予)。

此方法可讓您根據屬性控制存取，確保只有具有適當標籤的使用者或角色才能對指定的資源執行特定動作。

使用 ABAC 授予許可 (AWS CLI)

下列範例顯示必須滿足才能接收資源上所有可用許可的屬性表達式。或者，您可以指定個別許可`Select`，例如 `Describe`、 或 `Drop`。表達式使用 Cedar 政策表達式。如需 Cedar 的詳細資訊，請參閱什麼是 Cedar？| Cedar 政策語言參考GuideLink。

此條件會檢查 IAM 主體是否具有`department`標籤，且`department`標籤值是否等於 `sales`。

```
aws lakeformation grant-permissions
--principal '{"DataLakePrincipalIdentifier": "111122223333:IAMPrincipals"}'
--resource '{"Database": {"CatalogId": 111122223333, "Name": "abac-db"}}'
--permissions ALL
--condition '{"Expression": "context.iam.principalTags.hasTag(\"department\")
   && context.iam.principalTags.getTag(\"department\") == \"sales\""'
```

許可範例案例

下列案例有助於示範如何設定 許可，以安全存取 中的資料 AWS Lake Formation。

Shirley 是資料管理員。她想要為公司 AnyCompany 設定資料湖。目前，所有資料都存放在 Amazon S3 中。John 是行銷經理，需要客戶購買資訊的寫入存取權 （包含在 中`s3://customerPurchases`)。行銷分析師 Diego 今年夏天加入 John。John 需要能夠授予 Diego 存取權，以便在不涉及 Shirley 的情況下對資料執行查詢。

Mateo 需要財務部門的存取權來查詢會計資料 （例如 `s3://transactions`)。他想要查詢財務團隊使用的資料庫 (`Finance_DB`) 資料表中的交易資料。他的經理 Arnav 可以讓他存取 `Finance_DB`。雖然他不應該修改會計資料，但他需要將資料轉換為適合預測的格式 （結構描述）。此資料會存放在他可以修改的個別儲存貯體 (`s3://financeForecasts`) 中。

摘要：

- Shirley 是資料湖管理員。

- John 需要 CREATE_DATABASE和 CREATE_TABLE許可，才能在 Data Catalog 中建立新的資料庫和資料表。
- John 也需要 SELECT、 INSERT和 對其建立的資料表的DELETE許可。
- Diego 需要資料表的SELECT許可才能執行查詢。

AnyCompany 的員工會執行下列動作來設定許可。此案例中顯示的 API 操作會顯示簡化的語法，以求清晰。

1. Shirley 向 Lake Formation 註冊包含客戶購買資訊的 Amazon S3 路徑。

```
RegisterResource(ResourcePath("s3://customerPurchases"), false, Role_ARN )
```

2. Shirley 授予 John 存取包含客戶購買資訊的 Amazon S3 路徑。

```
GrantPermissions(John, S3Location("s3://customerPurchases"),
 [DATA_LOCATION_ACCESS]) )
```

3. Shirley 授予 John 建立資料庫的許可。

```
GrantPermissions(John, catalog, [CREATE_DATABASE])
```

4. John 建立資料庫 John_DB。John 會自動擁有該資料庫的CREATE_TABLE許可，因為他建立了該資料庫。

```
CreateDatabase(John_DB)
```

5. John 建立John_Table指向 的資料表s3://customerPurchases。因為他建立了資料表，所以他對其擁有所有許可，並且可以授予許可。

```
CreateTable(John_DB, John_Table)
```

6. John 允許他的分析師 Diego 存取資料表 John_Table。

```
 GrantPermissions(Diego, John_Table, [SELECT])
```

7. John 允許他的分析師 Diego 存取 s3://customerPurchases/London/。由於 Shirley 已註冊 s3://customerPurchases，因此其子資料夾會向 Lake Formation 註冊。

```
GrantDataLakePrivileges( 123456789012/datalake, Diego, [DATA_LOCATION_ACCESS], [],
S3Location("s3://customerPurchases/London/") )
```

8. John 允許他的分析師 Diego 在資料庫中建立資料表John_DB。

```
GrantDataLakePrivileges( 123456789012/datalake, Diego, John_DB, [CREATE_TABLE],
[] )
```

9. Diego John_DB 在 的 中建立資料表，s3://customerPurchases/London/並自動取得 ALTER、DROP、INSERT、 SELECT和 DELETE許可。

```
CreateTable( 123456789012/datalake, John_DB, Diego_Table )
```

Lake Formation 中的資料篩選和儲存格層級安全性

當您授予資料目錄資料表的 Lake Formation 許可時，您可以包含資料篩選規格，以限制對查詢結果和與 Lake Formation 整合的引擎中特定資料的存取。Lake Formation 使用資料篩選來實現資料欄層級安全性、資料列層級安全性和儲存格層級安全性。如果您的來源資料包含巢狀結構，您可以在巢狀資料欄上定義和套用資料篩選條件。

透過 Lake Formation 的資料篩選功能，您可以實作以下層級的資料安全。

資料欄層級安全

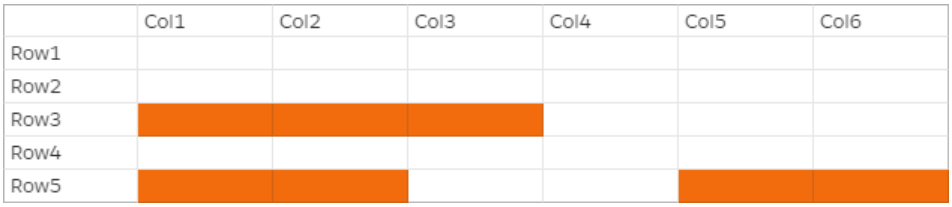
對資料目錄資料表授予具有資料欄層級安全性（資料欄篩選）的許可，可讓使用者僅檢視其在資料表中可存取的特定資料欄和巢狀資料欄。考慮在大型多區域通訊公司的多個應用程式中使用的persons資料表。授予資料目錄資料表的許可，若使用者不在 HR 部門工作，則無法查看個人身分識別資訊 (PII)，例如社會安全號碼或出生日期。您也可以定義安全政策，並僅授予對巢狀資料欄部分子結構的存取權。

資料列層級安全性

對資料目錄資料表授予具有資料列層級安全性（資料列篩選）的許可，可讓使用者僅檢視其在資料表中可存取的特定資料列。篩選是根據一或多個資料欄的值。您可以在定義資料列篩選條件表達式時包含巢狀資料欄結構。例如，如果通訊公司的不同區域辦公室有自己的人力資源部門，您可以將人力資源員工可以查看的人員記錄限制為僅限其區域中員工的記錄。

儲存格層級安全性

儲存格層級安全性結合了資料列篩選和資料欄篩選，可提供高度彈性的許可模型。如果您將資料表的資料列和資料欄檢視為網格，您可以使用儲存格層級安全性，限制存取網格的兩個維度的任何位置。也就是說，您可以根據資料列限制對不同資料欄的存取。下圖說明了此部分，其中限制的資料欄會著色。



繼續人員資料表的範例，您可以在儲存格層級建立資料篩選條件，如果資料列的國家/地區資料欄設定為 "UK"，則限制對街道地址資料欄的存取，但如果資料列的國家/地區資料欄設定為 "US"，則允許存取街道地址資料欄。

篩選條件僅適用於讀取操作。因此，您只能使用篩選條件授予 SELECT Lake Formation 許可。

巢狀資料欄上的儲存格層級安全性

Lake Formation 可讓您在巢狀資料欄上定義和套用具有儲存格層級安全性的資料篩選條件。不過，Amazon Athena、Amazon EMR 和 Amazon Redshift Spectrum 等整合分析引擎支援對具有資料列和資料欄層級安全性的 Lake Formation 受管巢狀資料表執行查詢。

如需限制的詳細資訊，請參閱資料篩選限制。

主題

- Lake Formation 中的資料篩選條件
- 資料列篩選條件表達式中的 PartiQL 支援
- 使用儲存格層級篩選查詢資料表所需的許可
- 管理資料篩選條件

Lake Formation 中的資料篩選條件

您可以建立資料篩選條件，來實作資料欄層級、資料列層級和儲存格層級安全性。當您授予資料表的 SELECT Lake Formation 許可時，請選取資料篩選條件。如果您的資料表包含巢狀資料欄結構，您可以透過包含或排除子資料欄來定義資料篩選條件，並在巢狀屬性上定義資料列層級篩選條件表達式。

每個資料篩選條件都屬於 Data Catalog 中的特定資料表。資料篩選條件包含下列資訊：

- 篩選器名稱

- 與篩選條件相關聯的資料表目錄 IDs
- 資料表名稱
- 包含資料表的資料庫名稱
- 資料欄規格 – 要包含或排除在查詢結果中的資料欄和巢狀資料欄 （含`struct`資料類型） 清單。
- 資料列篩選條件表達式 – 指定要包含在查詢結果中的資料列的表達式。在某些限制下，表達式具有 PartiQL 語言中 `WHERE`子句的語法。若要指定所有資料列，請在主控台中選擇存取資料列層級存取下的所有資料列，或在 API 呼叫`AllRowsWildcard`中使用 。

  如需資料列篩選條件表達式中支援哪些內容的詳細資訊，請參閱 資料列篩選條件表達式中的 PartiQL 支援。

您取得的篩選層級取決於資料篩選條件的填入方式。

- 如果您指定「所有資料行」萬用字元並提供資料列篩選條件運算式，則只會建立資料列層級安全性 (資料列篩選)。
- 當您包含或排除特定資料欄和巢狀資料欄，並使用全列萬用字元指定「所有資料列」時，您只會建立資料欄層級的安全性 （資料欄篩選）。
- 如果包含或排除特定資料行並提供資料列篩選條件運算式，則會建立儲存格層級安全性 (儲存格篩選)。

下列 Lake Formation 主控台的螢幕擷取畫面顯示執行儲存格層級篩選的資料篩選條件。對於資料表的查詢`orders`，它會限制對 `customer_name`欄的存取，而查詢結果只會傳回包含 'pharma' `product_type`欄的資料列。

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

請注意，使用單引號括住字串常值 'pharma'。

您可以使用 Lake Formation 主控台來建立此資料篩選條件，也可以將下列請求物件提供給 CreateDataCellsFilter API 操作。

```
{
     "Name": "restrict-pharma",
     "DatabaseName": "sales",
     "TableName": "orders",
     "TableCatalogId": "111122223333",
     "RowFilter": {"FilterExpression": "product_type='pharma'"},
     "ColumnWildcard": {
         "ExcludedColumnNames": ["customer_name"]
     }
}
```

您可以建立資料表所需的任意數量的資料篩選條件。若要這樣做，您需要資料表上具有授予選項的SELECT許可。根據預設，Data Lake 管理員具有在該帳戶中的所有資料表上建立資料篩選條件的許可。您通常只在將資料表的許可授予委託人時，才使用可能的資料篩選條件子集。例如，您可以為資料表建立第二個資料篩選條件，該orders資料表是row-security-only的資料篩選條件。請參閱上述螢幕擷取畫面，您可以選擇存取所有資料欄選項，並包含 的資料列篩選條件表達式product_type<>pharma。此資料篩選條件的名稱可以是 no-pharma。它會限制對資料product_type欄設定為 'pharma' 的所有資料列的存取。

此資料篩選條件的 CreateDataCellsFilter API 操作請求物件如下。

```
{
     "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"]
}
```

然後，您可以將orders資料表SELECT上的restrict-pharma資料篩選條件授予管理使用者，將orders資料表SELECT上的no-pharma資料篩選條件授予非管理使用者。對於醫療保健部門的使用者，您可以在orders資料表SELECT上授予 ，以完整存取所有資料列和資料欄 （無資料篩選條件），或者使用另一個限制對定價資訊的存取的資料篩選條件。

在資料篩選條件中指定資料欄層級和資料列層級安全性時，您可以包含或排除巢狀資料欄。在下列範例中，使用合格的資料欄名稱（以雙引號包裝）指定對 `product.offer` 欄位的存取。這對於巢狀欄位很重要，以避免資料欄名稱包含特殊字元時發生錯誤，並維持與最上層資料欄層級安全定義的回溯相容性。

```
{
     "Name": "example_dcf",
     "DatabaseName": "example_db",
     "TableName": "example_table",
     "TableCatalogId": "111122223333",
     "RowFilter": { "FilterExpression": "customer.customerName <> 'John'" },
     "ColumnNames": ["customer", "\"product\".\"offer\""]
}
```

另請參閱

- 管理資料篩選條件

資料列篩選條件表達式中的 PartiQL 支援

您可以使用 PartiQL 資料類型、運算子和彙總的子集來建構資料列篩選條件表達式。Lake Formation 不允許篩選表達式中的任何使用者定義或標準 partiQL 函數。您可以使用比較運算子來比較資料欄與常數（例如 `views >= 10000`），但無法比較資料欄與其他資料欄。

資料列篩選條件表達式可以是簡單表達式或複合表達式。表達式的總長度必須小於 2048 個字元。

簡單表達式

簡單表達式的格式如下： `<column name > <comparison operator ><value >`

- 資料欄名稱

  它可以是頂層資料欄、分割區欄或資料表結構描述中存在的巢狀資料欄，且必須屬於下列支援的資料類型內容。
- 比較運算子

  以下是支援的運算子： `=, >, <, >=, <=, <>,!=, BETWEEN, IN, LIKE, NOT, IS [NOT] NULL`

- 所有字串比較和LIKE模式比對都區分大小寫。您無法在分割區資料欄上使用 IS 【NOT】 NULL 運算子。
- 資料欄值

  資料欄值必須符合資料欄名稱的資料類型。

複合表達式

複合表達式的格式為：(<simple expression >) <AND/OR >(<simple expression >)。您可以使用邏輯運算子 進一步組合複合表達式AND/OR。

支援的資料類型

參考包含不支援資料類型之 AWS Glue Data Catalog 資料表的資料列篩選條件將導致錯誤。以下是資料表資料欄和常數支援的資料類型，這些資料類型會對應至 Amazon Redshift 資料類型：

- STRING, CHAR, VARCHAR
- INT, LONG, BIGINT, FLOAT, DECIMAL, DOUBLE
- BOOLEAN
- STRUCT

如需 Amazon Redshift 中資料類型的詳細資訊，請參閱《Amazon Redshift 資料庫開發人員指南》中的資料類型。

資料列篩選條件表達式

Example

以下是具有資料欄之資料表的有效資料列篩選條件表達式範例： 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

以下是具有巢狀資料欄之資料表的資料列篩選表達式的有效範例： `year > 2010 and customer.customerId <> 1`

定義巢狀資料列層級表達式時，不應參考分割區資料欄下的巢狀欄位。

字串常數必須以單引號括住。

保留的關鍵字

如果您的資料列篩選條件表達式包含 PartiQL 關鍵字，您會收到剖析錯誤，因為資料欄名稱可能與關鍵字衝突。發生這種情況時，請使用雙引號逸出資料欄名稱。預留關鍵字的一些範例包括「first」、「last」、「asc」、「missing」。如需預留關鍵字清單，請參閱 PartiQL 規格。

PartiQL 參考

如需 PartiQL 的詳細資訊，請參閱 https://partiql.org/。

使用儲存格層級篩選查詢資料表所需的許可

需要下列 AWS Identity and Access Management (IAM) 許可，才能對具有儲存格層級篩選的資料表執行查詢。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "lakeformation:StartQueryPlanning",
                "lakeformation:GetQueryState",
                "lakeformation:GetWorkUnits",
                "lakeformation:GetWorkUnitResults"
            ],
            "Resource": "*"
        }
    ]
}
```

如需 Lake Formation 許可的詳細資訊，請參閱 Lake Formation 角色和 IAM 許可參考。

管理資料篩選條件

若要實作資料欄層級、資料列層級和儲存格層級的安全性，您可以建立和維護資料篩選條件。每個資料篩選條件都屬於 Data Catalog 資料表。您可以為資料表建立多個資料篩選條件，然後在授予資料表許可時使用一或多個資料篩選條件。您也可以在具有`struct`資料類型的巢狀資料欄上定義和套用資料篩選條件，讓使用者只能存取巢狀資料欄的子結構。

您需要具有授予選項的`SELECT`許可，才能建立或檢視資料篩選條件。若要允許帳戶中的主體檢視和使用資料篩選條件，您可以授予其`DESCRIBE`許可。

> **Note**
>
> Lake Formation 不支援授予從另一個帳戶共用的資料篩選條件`Describe`許可。

您可以使用 AWS Lake Formation 主控台、 API 或 () AWS Command Line Interface 來管理資料篩選條件AWS CLI。

如需資料篩選條件的資訊，請參閱 Lake Formation 中的資料篩選條件

建立資料篩選條件

您可以為每個 Data Catalog 資料表建立一或多個資料篩選條件。

建立 Data Catalog 資料表的資料篩選條件 （主控台）

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

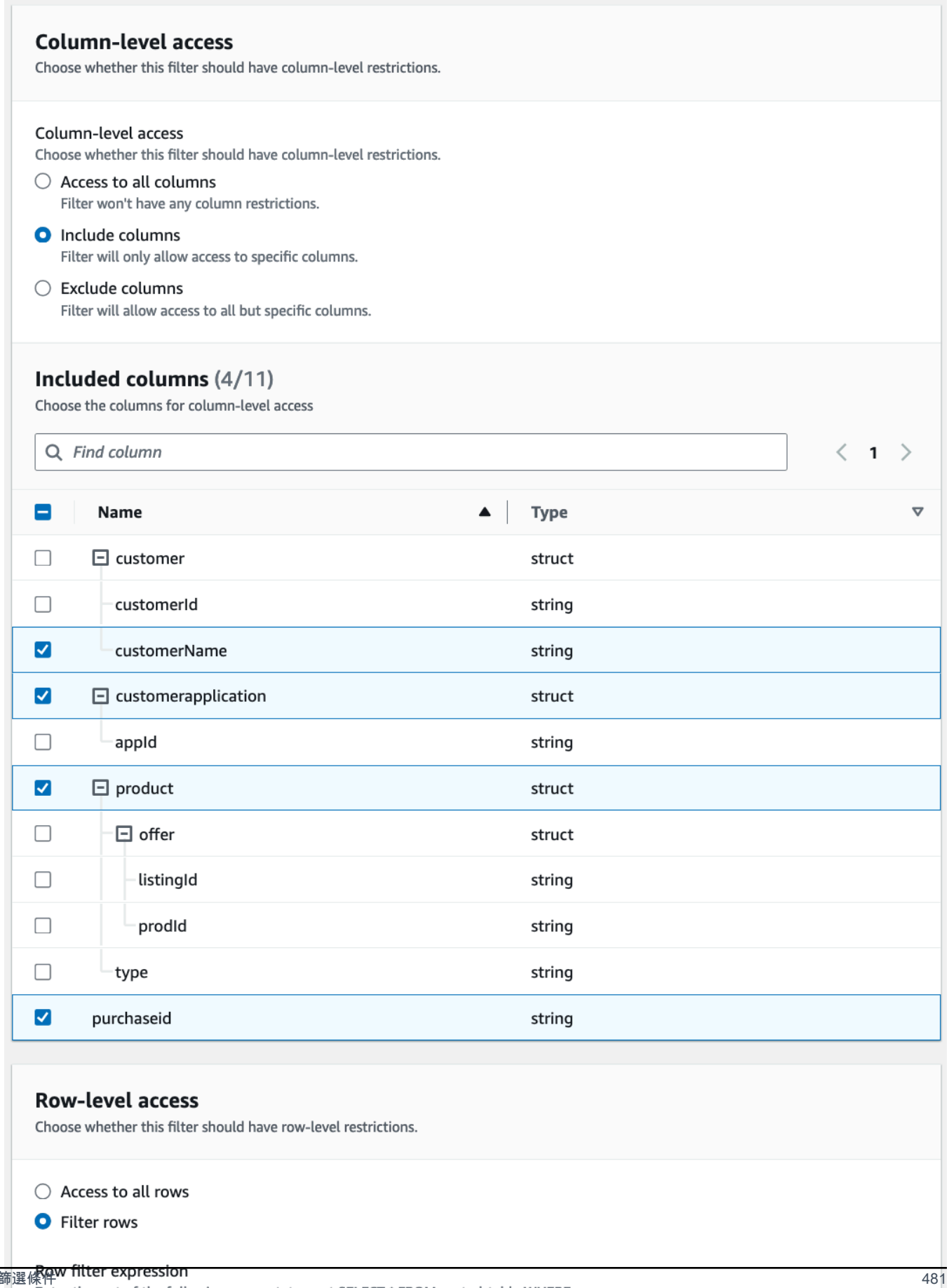
   以資料湖管理員、目標資料表擁有者或擁有目標資料表 Lake Formation 許可的主體身分簽署。
2. 在導覽窗格中的資料目錄下，選擇資料篩選條件。
3. 在資料篩選條件頁面上，選擇建立新篩選條件。
4. 在建立資料篩選條件對話方塊中，輸入下列資訊：

   - 資料篩選條件名稱
   - 目標資料庫 – 指定包含資料表的資料庫。
   - 目標資料表
   - 資料欄層級存取 – 將此設定為存取所有資料欄，以僅指定資料列篩選。選擇包含資料欄或排除資料欄以指定資料欄或儲存格篩選，然後指定要包含或排除的資料欄。

巢狀資料欄 – 如果您要在包含巢狀資料欄的資料表上套用篩選條件，您可以明確指定資料篩選條件中巢狀結構資料欄的子結構。

當您將 SELECT 許可授予此檔案伺服器上的委託人時，執行下列查詢的委託人只會看到 的資料，`customer.customerName`而不會看到 的資料`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'

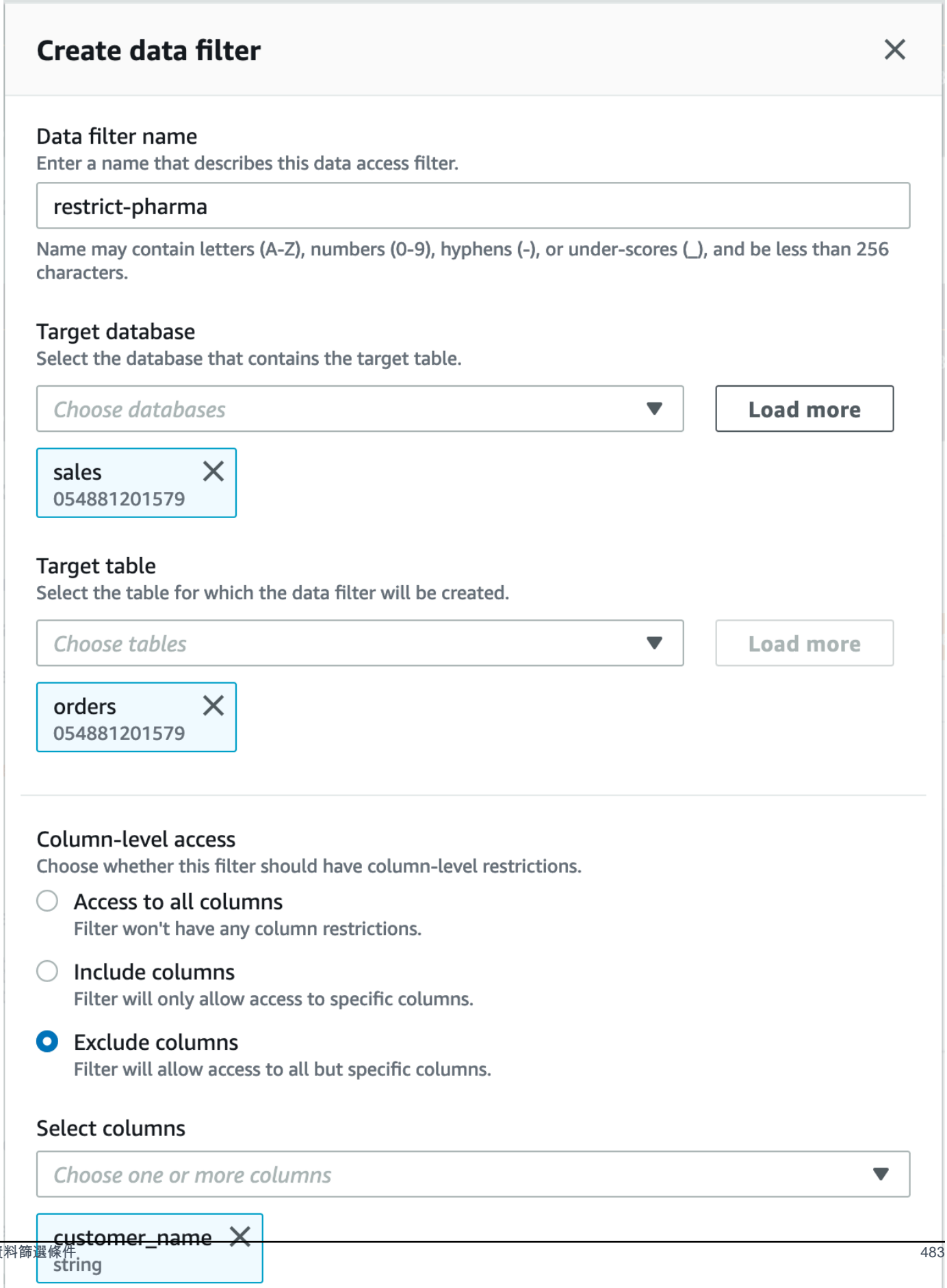
當您授予資料customer欄的許可時，委託人會收到資料欄的存取權，以及資料欄 (customerName 和) 下的巢狀欄位customerID。

- 資料列篩選條件表達式 – 輸入篩選條件表達式以指定資料列或儲存格篩選。如需支援的資料類型和運算子，請參閱 資料列篩選條件表達式中的 PartiQL 支援。選擇存取所有資料列，以授予存取所有 的權限。

  您可以在資料列篩選條件表達式中包含巢狀資料欄的部分資料欄結構，以篩選包含特定值的資料列。

  當主體被授予具有資料列篩選條件表達式 的資料表許可Select * from example_nestedtable where customer.customerName <>'John'，且資料欄層級存取設定為存取所有資料欄時，查詢結果只會顯示 customerName <>'John'評估為 true 的資料列。

下列螢幕擷取畫面顯示實作儲存格篩選的資料篩選條件。在對orders資料表的查詢中，它會拒絕存取資料customer_name欄，並只顯示資料product_type欄中具有「製藥」的資料列。


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. 選擇 Create filter (建立篩選條件)。

在巢狀欄位上使用儲存格篩選條件政策建立資料篩選條件

本節使用以下範例結構描述，示範如何建立資料儲存格篩選條件：

```
[
    { 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. 在建立資料篩選條件上，頁面輸入資料篩選條件的名稱。
2. 接著，使用下拉式清單選擇資料庫名稱和資料表名稱。
3. 在資料欄層級存取區段中，選擇包含的資料欄，然後選取巢狀資料欄 (`customer.customerName`)。
4. 在資料列層級存取區段中，選擇存取所有資料列選項。
5. 選擇 Create filter (建立篩選條件)。

   當您授予此篩選條件的`SELECT`許可時，委託人可以存取 `customerName`資料欄中的所有資料列。
6. 接著，為相同的資料庫/資料表定義另一個資料篩選條件。
7. 在資料欄層級存取區段中，選擇包含的資料欄，然後選取另一個巢狀資料欄 (`customer.customerid`)。
8. 在資料列層級存取區段中，選擇篩選資料列，然後輸入資料列篩選表達式 (`customer.customerid <> 5`)。
9. 選擇 Create filter (建立篩選條件)。

   當您授予此篩選條件的`SELECT`許可時，委託人會收到 中所有資料列的存取權`customerName`，以及資料`customerId`欄中值為 5 的儲存格以外的`customerId`欄位。

授予資料篩選條件許可

您可以將資料篩選條件的 `SELECT`、 `DESCRIBE`和 `DROP` Lake Formation 許可授予委託人。

一開始，只有您可以檢視為資料表建立的資料篩選條件。若要讓其他主體檢視資料篩選條件，並授予資料目錄許可，您必須：

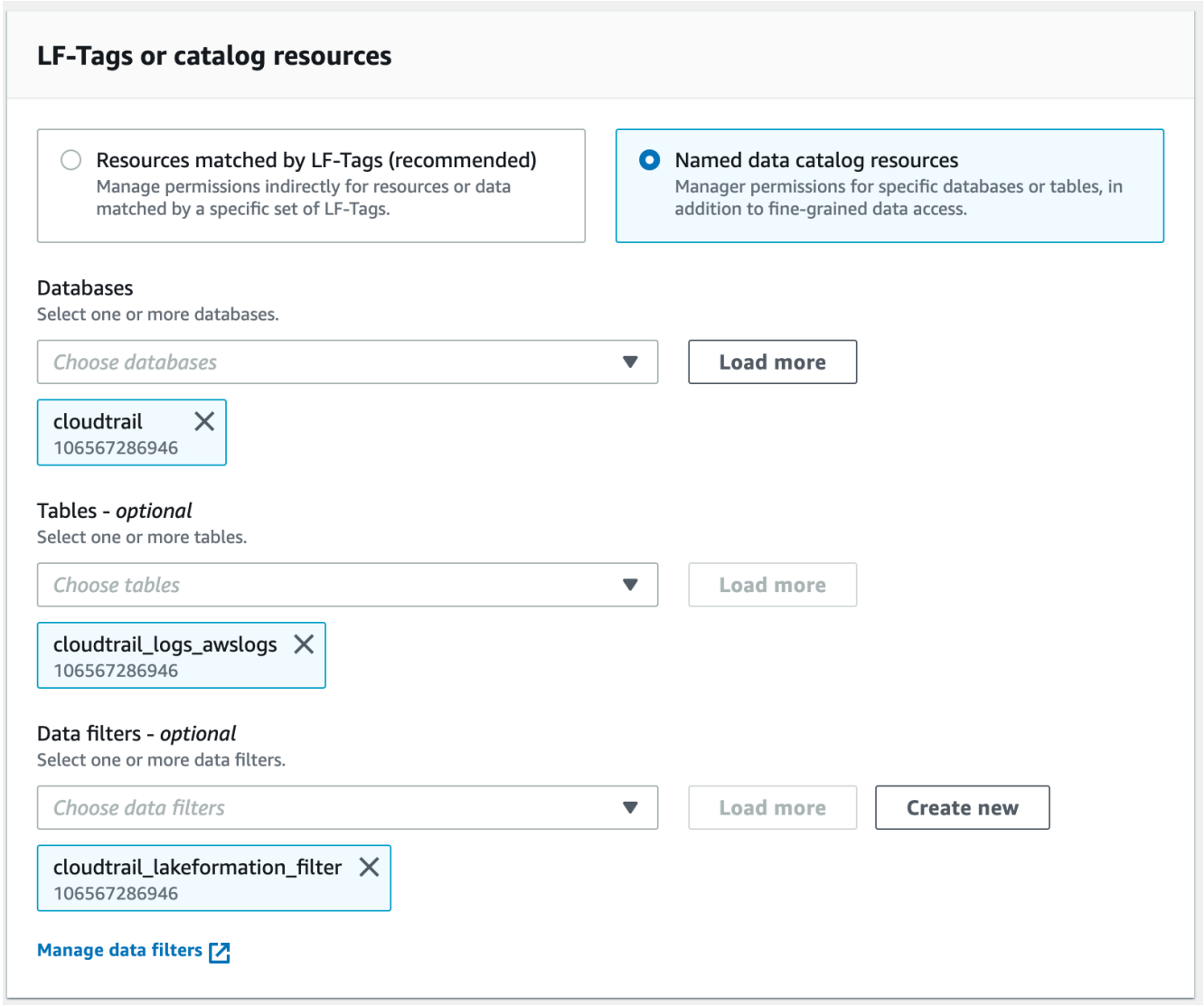
- 使用授予選項將資料表SELECT上的 授予委託人，並將資料篩選條件套用至授予。
- 將資料篩選條件的 DESCRIBE或 DROP許可授予委託人。

您可以將 SELECT許可授予外部 AWS 帳戶。然後，該帳戶中的資料湖管理員可以將該許可授予帳戶中的其他主體。授予外部帳戶時，您必須包含授予選項，外部帳戶的管理員才能進一步將許可串聯至其帳戶中的其他使用者。授予您帳戶中的委託人時，使用授予選項進行授予是選用的。

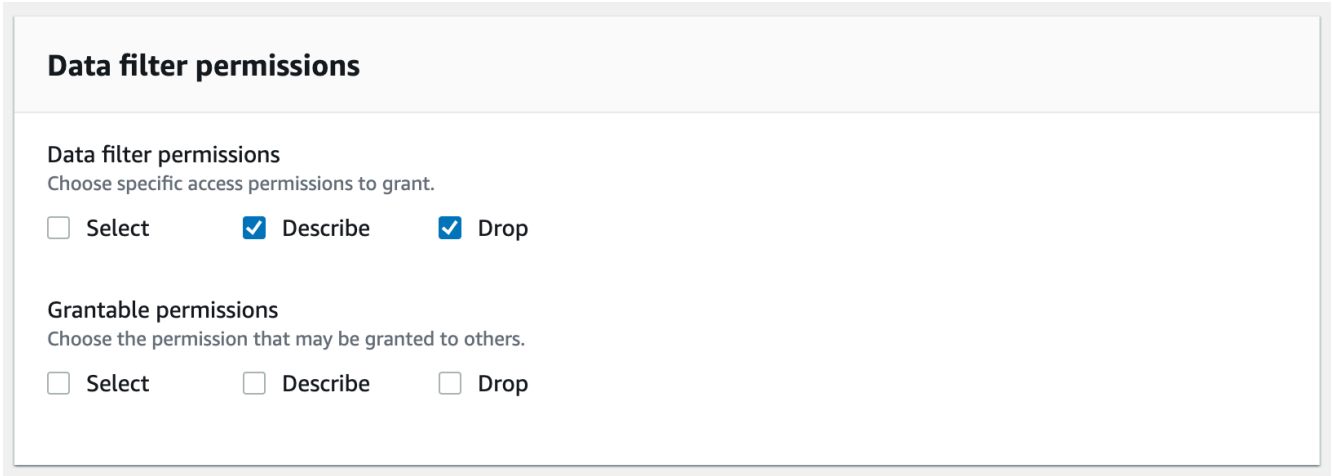
您可以使用 AWS Lake Formation 主控台、 API 或 AWS Command Line Interface () 來授予和撤銷資料篩選條件的許可AWS CLI。

Console

1. 登入 AWS Management Console 並開啟 Lake Formation 主控台，網址為 https：//https://console.aws.amazon.com/lakeformation/.。
2. 在導覽窗格的許可下，選擇資料湖許可。
3. 在許可頁面上的資料許可區段中，選擇授予。
4. 在授予資料許可頁面上，選擇要授予許可的委託人。
5. 在 LF 標籤或目錄資源區段中，選擇具名資料目錄資源。然後選擇您要授予許可的資料庫、資料表和資料篩選條件。



6. 在資料篩選條件許可區段中，選擇您要授予所選委託人的許可。



AWS CLI

- 輸入grant-permissions命令。DataCellsFilter 為resource引數指定 ，並為Permissions引數指定 DROP DESCRIBE或 ，並選擇性地為PermissionsWithGrantOption引數指定 或 。

  下列範例DESCRIBE使用 授予選項授予資料篩選條件 datalake_user1上的使用者restrict-pharma，該選項屬於 AWS 帳戶 1111-2222-3333 中sales資料庫中的 orders 資料表。

```
aws lakeformation grant-permissions --cli-input-json file://grant-params.json
```

  以下是檔案 的內容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"]
}
```

授予資料篩選條件提供的資料許可

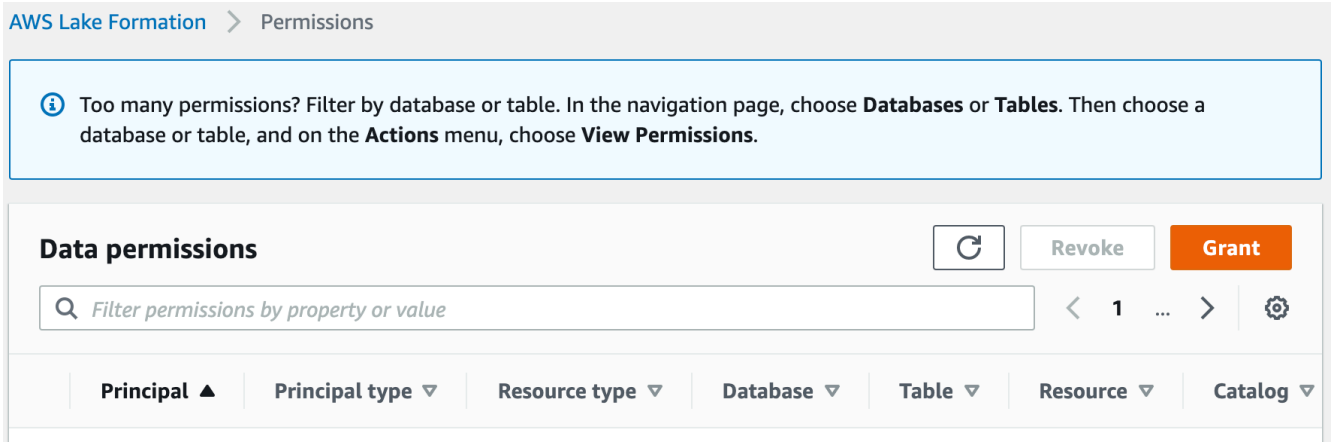
資料篩選條件代表資料表中的資料子集。若要提供資料存取權給委託人，需要將SELECT許可授予這些委託人。透過此許可，委託人可以：

- 在與其帳戶共用的資料表清單中檢視實際的資料表名稱。
- 在共用資料表上建立資料篩選條件，並在這些資料篩選條件上授予使用者許可。

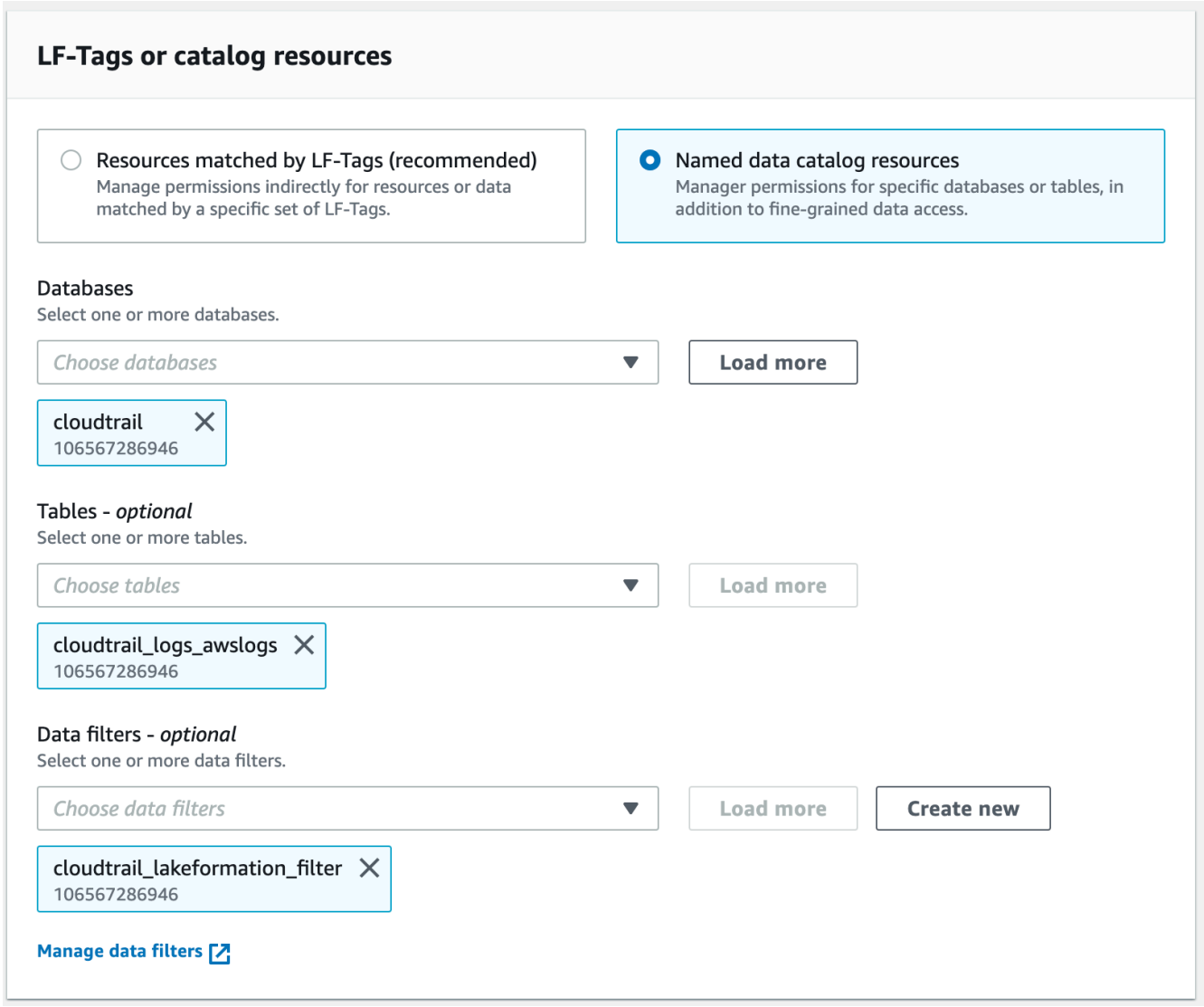
Console

授予 SELECT 許可

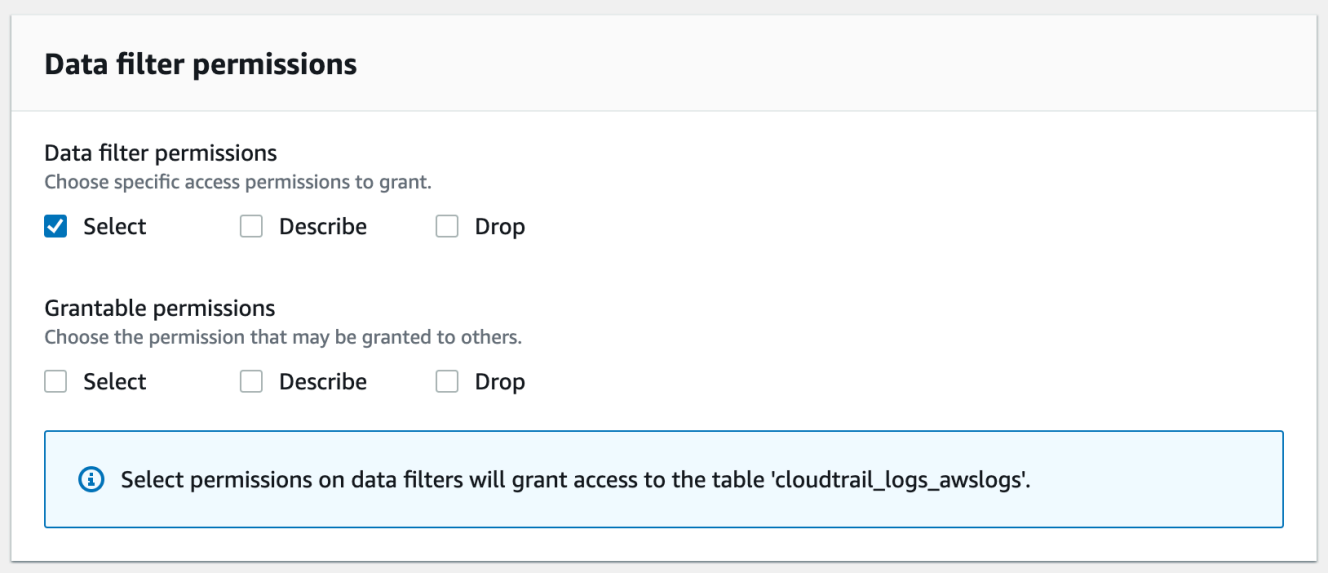
1. 前往 Lake Formation 主控台中的許可頁面，然後選擇授予。



2. 選取您要提供存取權的主體，然後選取具名資料目錄資源。



3. 若要提供篩選條件所代表資料的存取權，請選擇資料篩選條件許可下的選取。



CLI

輸入grant-permissions命令。DataCellsFilter 為資源引數指定 ，並為SELECT許可引數指定 。

下列範例SELECT使用 授予選項授予資料篩選條件 datalake_user1上的使用者restrict-pharma，該選項屬於 中sales資料庫中的 orders資料表 AWS 帳戶 1111-2222-3333。

```
aws lakeformation grant-permissions --cli-input-json file://grant-params.json
```

以下是檔案 的內容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"]
```

```
}
```

檢視資料篩選條件

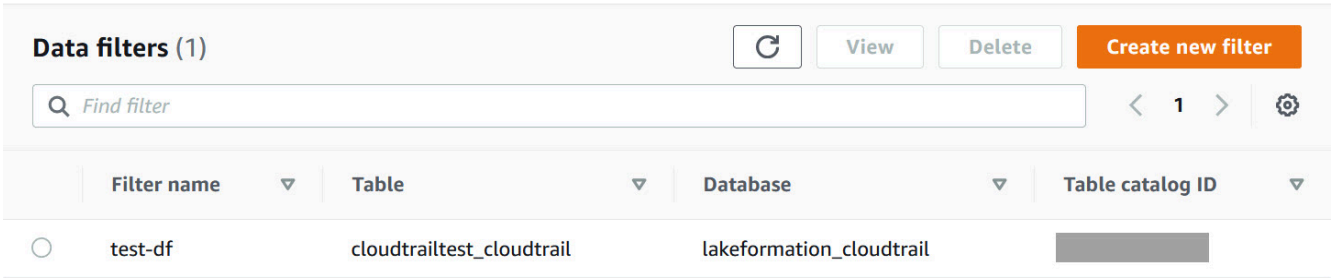
您可以使用 Lake Formation 主控台 AWS CLI或 Lake Formation API 來檢視資料篩選條件。

若要檢視資料篩選條件，您必須是 Data Lake 管理員或具有資料篩選條件所需的許可。

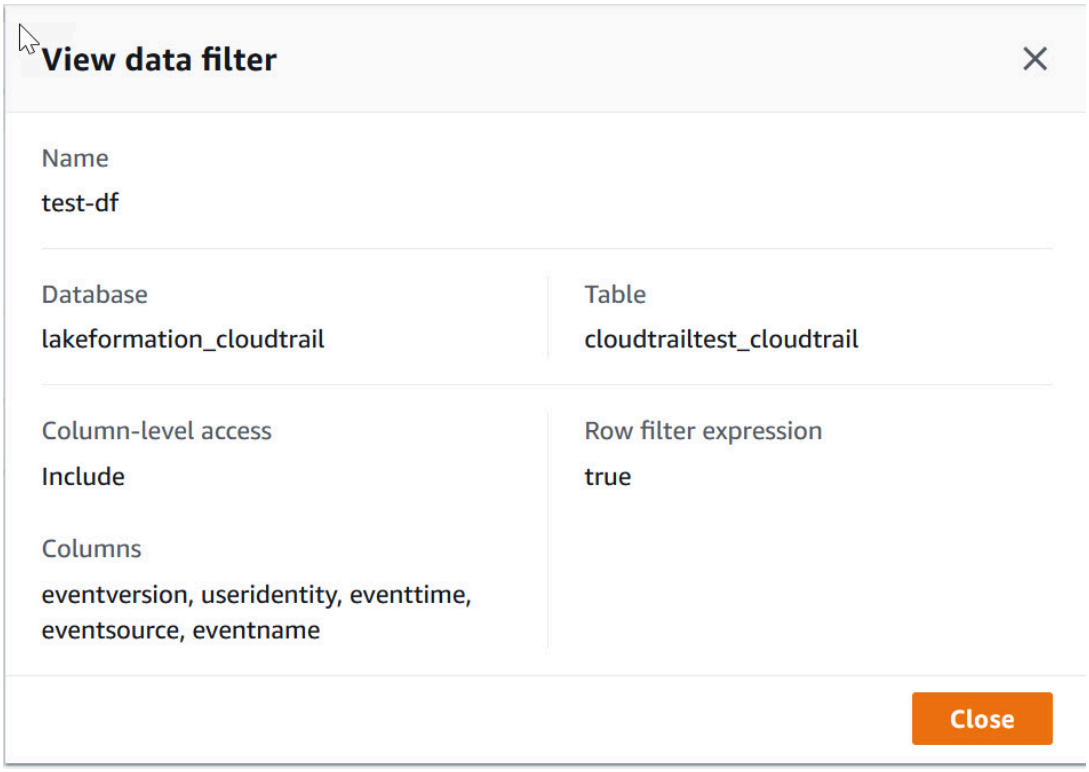
Console

1. 登入 AWS Management Console 並開啟 Lake Formation 主控台，網址為 https：/https://console.aws.amazon.com/lakeformation/.。
2. 在導覽窗格中的資料目錄下，選擇資料篩選條件。

   此頁面會顯示您有權存取的資料篩選條件。



3. 若要檢視資料篩選條件詳細資訊，請選擇資料篩選條件，然後選擇檢視。隨即出現新視窗，其中包含資料篩選條件詳細資訊。



AWS CLI

輸入`list-data-cells-filter`命令並指定資料表資源。

下列範例列出`cloudtrailtest_cloudtrail`資料表的資料篩選條件。

```
aws lakeformation list-data-cells-filter --table '{ "CatalogId":"123456789012",
"DatabaseName":"lakeformation_cloudtrail", "Name":"cloudtrailtest_cloudtrail"}'
```

API/SDK

使用 `ListDataCellsFilter` API 並指定資料表資源。

下列範例使用 Python 列出`myTable`資料表的前 20 個資料篩選條件。

```
response = client.list_data_cells_filter(
    Table = {
        'CatalogId': '111122223333',
        'DatabaseName': 'mydb',
        'Name': 'myTable'
    },
    MaxResults=20
```

```
)
```

列出資料篩選條件許可

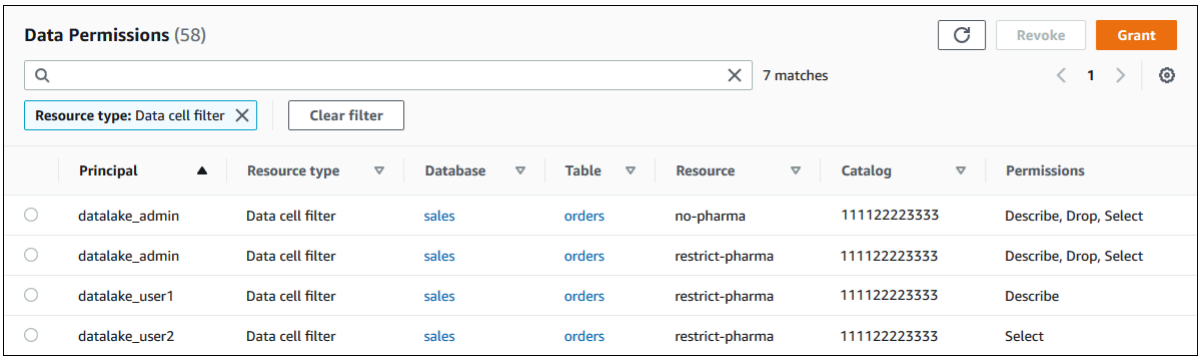
您可以使用 Lake Formation 主控台來檢視資料篩選條件上授予的許可。

若要檢視資料篩選條件的許可，您必須是 Data Lake 管理員或具有資料篩選條件所需的許可。

Console

1. 登入 AWS Management Console 並開啟 Lake Formation 主控台，網址為 https：//https://console.aws.amazon.com/lakeformation/.。
2. 在導覽窗格的許可下，選擇資料許可。
3. 在資料許可頁面上，按一下或點選搜尋欄位，然後在屬性功能表中選擇資源類型。
4. 在資源類型功能表中，選擇資源類型：資料儲存格篩選條件。

   列出您擁有許可的資料篩選條件。您可能需要水平捲動，才能查看許可和可授予資料欄。



AWS CLI

- 輸入`list-permissions`命令。`DataCellsFilter` 為`resource`引數指定 ，並為`Permissions`引數指定 `DROP` `DESCRIBE`或 ，並選擇性地為`PermissionsWithGrantOption`引數指定 `SELECT` 或 `ALL`。

  下列範例會列出資料篩選條件 上具有授予選項的`DESCRIBE`許可`restrict-pharma`。結果僅限於授予 主體的許可，`datalake_user1`以及 AWS 帳戶 1111-2222-3333 中`sales`資料庫中的`orders`資料表。

```
aws lakeformation list-permissions --cli-input-json file://list-params.json
```

以下是檔案 的內容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"]
}
```

在 Lake Formation 中檢視資料庫和資料表許可

您可以檢視在 Data Catalog 資料庫或資料表上授予的 Lake Formation 許可。您可以使用 Lake Formation 主控台、API 或 AWS Command Line Interface () 來執行此操作AWS CLI。

使用 主控台，您可以檢視從資料庫或資料表頁面或資料許可頁面開始的許可。

> **Note**
>
> 如果您不是資料庫管理員或資源擁有者，則只有在您對具有授予選項的資源具有 Lake Formation 許可時，才能檢視其他主體在資源上擁有的許可。
> 除了必要的 Lake Formation 許可之外，您還需要 AWS Identity and Access Management (IAM) 許可 `glue:GetDatabases`、`glue:GetDatabase`、`glue:GetTables`、`glue:GetTable`和 `lakeformarion:ListPermissions`。

檢視資料庫的許可 （ 主控台，從資料庫頁面開始 ）

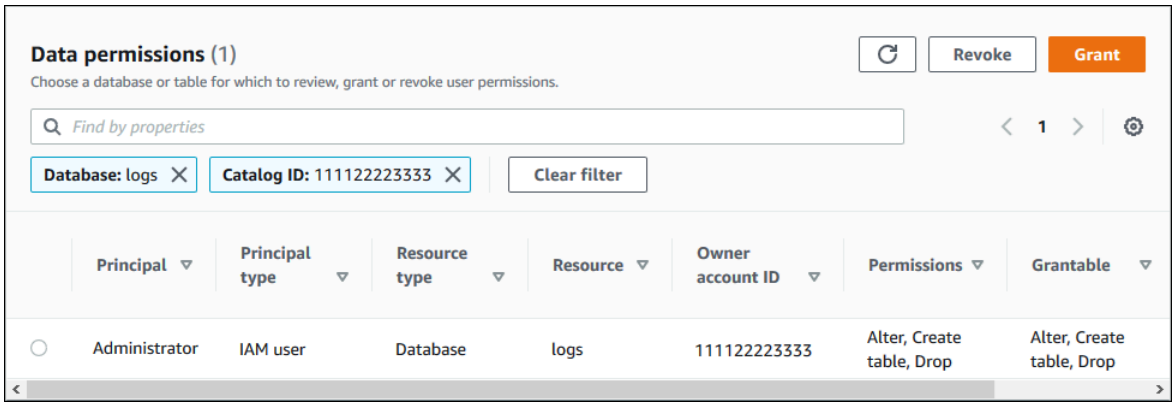
1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

   以資料湖管理員、資料庫建立者或具有資料庫上任何 Lake Formation 許可的使用者身分登入，並使用 授予選項。

2. 在導覽窗格中，選擇 Databases (資料庫)。
3. 選擇資料庫，然後在動作功能表上，選擇檢視許可。

> **Note**
>
> 如果您選擇資料庫資源連結，Lake Formation 會在資源連結上顯示許可，而不是資源連結的目標資料庫。

資料許可頁面列出資料庫的所有 Lake Formation 許可。資料庫擁有者的資料庫名稱和目錄 ID (AWS 帳戶 ID) 會顯示為搜尋方塊下的標籤。圖磚表示篩選條件已套用至僅列出該資料庫的許可。您可以關閉圖磚或選擇清除篩選條件來調整篩選條件。



檢視資料庫上的許可 （主控台，從資料許可頁面開始）

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

   以資料湖管理員、資料庫建立者或具有資料庫上任何 Lake Formation 許可的使用者身分登入，並使用 授予選項。
2. 在導覽窗格中，選擇資料許可。
3. 將游標放在頁面頂端的搜尋方塊中，然後在出現的屬性功能表中，選擇資料庫。
4. 在出現的資料庫功能表中，選擇資料庫。

> **Note**
>
> 如果您選擇資料庫資源連結，Lake Formation 會在資源連結上顯示許可，而不是資源連結的目標資料庫。

資料許可頁面列出資料庫的所有 Lake Formation 許可。資料庫名稱會顯示為搜尋方塊下的圖磚。圖磚表示篩選條件已套用至僅列出該資料庫的許可。您可以關閉圖磚或選擇清除篩選條件來移除篩選條件。

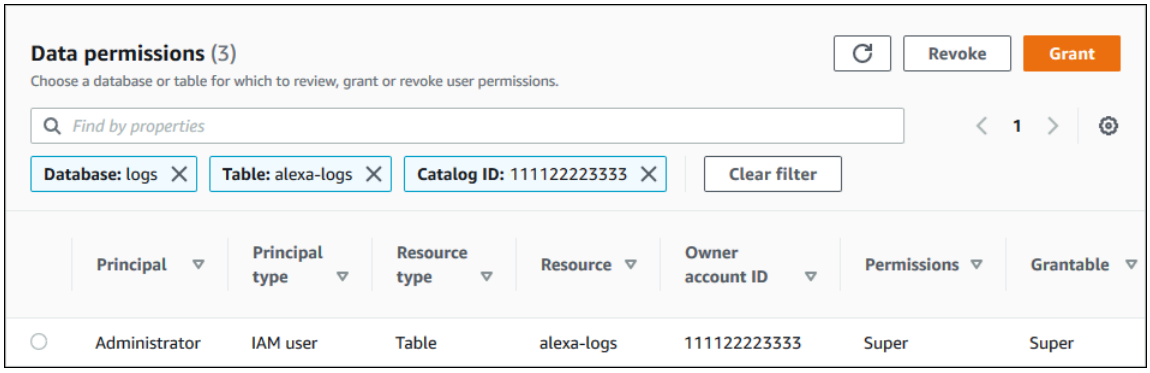
檢視資料表上的許可 （主控台，從資料表頁面開始）

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

   使用授予選項，以資料湖管理員、資料表建立者或擁有資料表上任何 Lake Formation 許可的使用者身分登入。
2. 在導覽窗格中，選擇 Tables (資料表)。
3. 選擇資料表，然後在動作功能表中選擇檢視許可。

> **Note**
>
> 如果您選擇資料表資源連結，Lake Formation 會在資源連結上顯示許可，而不是在資源連結的目標資料表上顯示許可。

資料許可頁面列出資料表的所有 Lake Formation 許可。資料表名稱、包含資料表的資料庫名稱，以及資料表擁有者的目錄 ID (AWS 帳戶 ID) 會顯示為搜尋方塊下的標籤。標籤指出篩選條件已套用至僅列出該資料表的許可。您可以關閉標籤或選擇清除篩選條件來調整篩選條件。



檢視資料表上的許可 （主控台，從資料許可頁面開始）

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

使用授予選項，以資料湖管理員、資料表建立者或擁有資料表上任何 Lake Formation 許可的使用者身分登入。

2. 在導覽窗格中，選擇資料許可。
3. 將游標放在頁面頂端的搜尋方塊中，然後在出現的屬性功能表中，選擇資料庫。
4. 在出現的資料庫功能表中，選擇資料庫。

> **Important**
>
> 如果您想要檢視從外部 AWS 帳戶與您的帳戶共用之資料表的許可，則必須選擇外部帳戶中包含資料表的資料庫，而不是資料庫的資源連結。

資料許可頁面列出資料庫的所有 Lake Formation 許可。

5. 再次將游標放置在搜尋方塊中，然後在出現的屬性功能表上，選擇資料表。
6. 在出現的資料表功能表上，選擇資料表。

資料許可頁面會列出資料表的所有 Lake Formation 許可。包含資料表之資料庫的資料表名稱和資料庫名稱會在搜尋方塊下顯示為圖磚。圖磚表示篩選條件已套用至僅列出該資料表的許可。您可以關閉圖磚或選擇清除篩選條件來調整篩選條件。

檢視資料表上的許可 (AWS CLI)

- 輸入`list-permissions`命令。

下列範例列出從外部帳戶共用之資料表的許可。`CatalogId` 屬性是外部 AWS 帳戶的帳戶 ID，資料庫名稱是指外部帳戶中包含資料表的資料庫。

```
aws lakeformation list-permissions  --resource-type TABLE --resource '{ "Table":
 {"DatabaseName":"logs", "Name":"alexa-logs", "CatalogId":"123456789012"}}'
```

使用 Lake Formation 主控台撤銷許可

您可以使用 主控台撤銷所有類型的 Lake Formation 許可：Data Catalog 許可、政策標籤許可、資料篩選條件許可和位置許可。

撤銷資源的 Lake Formation 許可（主控台）

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

   以資料湖管理員或使用者身分登入，該使用者已獲得資源上的授予選項的許可。
2. 在導覽窗格的許可下，選擇資料湖許可、LF 標籤和許可或資料位置。
3. 選取許可或位置，然後選擇撤銷。
4. 在開啟的對話方塊中，選擇撤銷。

Lake Formation 中的跨帳戶資料共用

Lake Formation 跨帳戶功能可讓使用者安全地跨多個 AWS 組織或直接與另一個帳戶中的 IAM 主體共用分散式資料湖 AWS 帳戶，提供對 Data Catalog 中繼資料和基礎資料的精細存取。大型企業通常會使用多個 AWS 帳戶，其中許多帳戶可能需要存取由單一 管理的資料湖 AWS 帳戶。使用者和 AWS Glue 擷取、轉換和載入 (ETL) 任務可以跨多個帳戶查詢和聯結資料表，但仍會利用 Lake Formation 資料表層級和資料欄層級的資料保護。

當您將 Data Catalog 資源的 Lake Formation 許可授予外部帳戶或直接授予另一個帳戶中的 IAM 主體時，Lake Formation 會使用 AWS Resource Access Manager (AWS RAM) 服務來共用資源。如果承授者帳戶與承授者帳戶位於相同的組織中，則承授者可以立即使用共用資源。如果承授者帳戶不在同一個組織中， AWS RAM 會傳送邀請給承授者帳戶，以接受或拒絕資源授予。然後，若要提供共用資源，承授者帳戶中的資料湖管理員必須使用 AWS RAM 主控台或 AWS CLI 接受邀請。

Lake Formation 支援以混合存取模式與外部帳戶共用 Data Catalog 資源。混合存取模式可讓您靈活地選擇性地為 中的資料庫和資料表啟用 Lake Formation 許可 AWS Glue Data Catalog。
使用混合存取模式時，您現在有一個增量路徑，可讓您為一組特定使用者設定 Lake Formation 許可，而不會中斷其他現有使用者或工作負載的許可政策。

如需詳細資訊，請參閱混合存取模式。

直接跨帳戶共用

授權委託人可以與外部帳戶中的 IAM 委託人明確共用資源。當帳戶擁有者想要控制外部帳戶中的誰可以存取資源時，此功能非常有用。IAM 主體收到的許可將是直接授予的聯集，而帳戶層級授予會層疊到主體。只有許可授予的收件人可以檢視直接跨帳戶授予。接收資源共用的委託人無法與其他委託人共用資源。

共用 Data Catalog 資源的方法

透過單一 Lake Formation 授予操作，您可以授予下列 Data Catalog 資源的跨帳戶許可。

- 資料庫
- 個別資料表（具有選用的資料欄篩選）
- 幾個選取的資料表
- 資料庫中的所有資料表（使用所有資料表萬用字元）

有兩個選項可讓您與另一個帳戶中的另一個 AWS 帳戶 或 IAM 主體共用資料庫和資料表。

- Lake Formation 標籤型存取控制 (LF-TBAC)（建議）

  Lake Formation 標籤型存取控制是一種根據屬性定義許可的授權策略。您可以使用標籤型存取控制，與外部 IAM 主體、組織和組織單位 (OUs) 共用 Data Catalog 資源（資料庫 AWS 帳戶、資料表和資料欄）。在 Lake Formation 中，這些屬性稱為 LF 標籤。如需詳細資訊，請參閱使用 Lake Formation 標籤型存取控制管理資料湖。

  > **Note**
  >
  > 將 Data Catalog 許可 AWS Resource Access Manager 用於跨帳戶授予的 LF-TBAC 方法。Lake Formation 現在支援使用 LF-TBAC 方法將跨帳戶許可授予 Organizations 和組織單位。
  > 若要啟用此功能，您需要將跨帳戶版本設定更新為第 3 版或更新版本。
  > 如需詳細資訊，請參閱更新跨帳戶資料共用版本設定。

- Lake Formation 命名資源

  使用具名資源方法的 Lake Formation 跨帳戶資料共用可讓您使用 Data Catalog 資料表和資料庫上的授予選項，將 Lake Formation 許可授予外部 AWS 帳戶、IAM 主體、組織或組織單位。授予操作會自動共用這些資源。

> **Note**
>
> 您也可以允許 AWS Glue 爬蟲程式使用 Lake Formation 登入資料來存取不同帳戶中的資料存放區。如需詳細資訊，請參閱《AWS Glue 開發人員指南》中的跨帳戶爬取。

Athena 和 Amazon Redshift Spectrum 等整合服務需要資源連結，才能在查詢中包含共用資源。如需資源連結的詳細資訊，請參閱 資源連結在 Lake Formation 中如何運作。

如需考量和限制，請參閱 跨帳戶資料共用最佳實務和考量事項。

主題

- 先決條件
- 更新跨帳戶資料共用版本設定
- 從外部帳戶跨 AWS 帳戶 或 IAM 主體共用 Data Catalog 資料表和資料庫
- 授予與您帳戶共用之資料庫或資料表的許可
- 授予資源連結許可
- 存取共用資料表的基礎資料
- 跨帳戶 CloudTrail 記錄
- 使用 AWS Glue和 Lake Formation 管理跨帳戶許可
- 使用 GetResourceShares API 操作檢視所有跨帳戶授予

相關主題

- Lake Formation 許可概觀
- 存取和檢視共用的 Data Catalog 資料表和資料庫
- 建立資源連結
- 對跨帳戶存取進行故障診斷

先決條件

在 AWS 您的帳戶可以與另一個帳戶或另一個帳戶中的委託人共用 Data Catalog 資源 （目錄、資料庫和資料表） 之前，以及在您可以存取與帳戶共用的資源之前，必須符合下列先決條件。

一般跨帳戶資料共用需求

- 若要在混合存取模式中共用 Data Catalog 資料庫和資料表，並在聯合目錄中共用物件，您需要將跨帳戶版本設定更新為第 4 版。
- 在授予 Data Catalog 資源的跨帳戶許可之前，您必須撤銷資源`IAMAllowedPrincipals`群組中的所有 Lake Formation 許可。如果呼叫委託人具有跨帳戶存取資源的許可，

且該IAMAllowedPrincipals許可存在於資源上，則 Lake Formation 會擲出AccessDeniedException。

此要求僅適用於在 Lake Formation 模式中註冊基礎資料位置時。如果您以混合模式註冊資料位置，IAMAllowedPrincipals群組許可可以存在於共用資料庫或資料表上。

- 對於包含您要共用之資料表的資料庫，您必須防止新資料表預設授與 Super給IAMAllowedPrincipals。在 Lake Formation 主控台上，編輯資料庫並關閉僅針對此資料庫中的新資料表使用 IAM 存取控制，或輸入下列 AWS CLI 命令，database將 取代為資料庫的名稱。如果基礎資料位置是在混合存取模式中註冊，您不需要變更此預設設定。在混合存取模式中，Lake Formation 可讓您在相同的資源 AWS Glue 上選擇性地強制執行 Amazon S3 和 的 Lake Formation 許可和 IAM 許可政策。

```
aws glue update-database --name database --database-input
 '{"Name":"database","CreateTableDefaultPermissions":[]}'
```

- 若要授予跨帳戶許可，授予者必須擁有 AWS Glue和 AWS RAM 服務所需的 AWS Identity and Access Management (IAM) 許可。 AWS 受管政策會AWSLakeFormationCrossAccountManager授予必要的許可。

使用 接收資源共享之帳戶中的資料湖管理員 AWS RAM 必須具有下列額外政策。它允許管理員接受 AWS RAM 資源共用邀請。它還允許管理員啟用與組織的資源共用。

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

- 如果您想要與 AWS Organizations 或組織單位共用 Data Catalog 資源，則必須在其中啟用與組織共用 AWS RAM。

  如需有關如何啟用與組織共用的資訊，請參閱AWS RAM 《 使用者指南》中的啟用與 AWS 組織共用。

  您必須擁有 `ram:EnableSharingWithAwsOrganization`許可，才能與組織共用。
- 若要直接與另一個帳戶中的 IAM 主體共用資源，您需要將跨帳戶版本設定更新為第 3 版。此設定可在資料目錄設定頁面上取得。如果您使用的是第 1 版，請參閱更新設定 的說明更新跨帳戶資料共用版本設定。
- 您無法與另一個帳戶共用使用 AWS Glue 服務受管金鑰加密的 Data Catalog 資源。您只能共用使用客戶加密金鑰加密的 Data Catalog 資源，而且接收資源共用的帳戶必須具有 Data Catalog 加密金鑰的許可，才能解密物件。

使用 LF-TBAC 需求的跨帳戶資料共用

- 若要與 AWS Organizations 和組織單位 (OUs) 共用 Data Catalog 資源，您需要將跨帳戶版本設定更新為版本 3 或更高版本。
- 若要與跨帳戶版本設定第 3 版共用 Data Catalog 資源，授予者需要擁有`AWSLakeFormationCrossAccountManager`您帳戶中 AWS 受管政策中定義的 IAM 許可。
- 如果您使用的是跨帳戶版本設定的第 1 版或第 2 版，您必須擁有啟用 LF-TBAC 的資料目錄資源政策 (`glue:PutResourcePolicy`)。如需詳細資訊，請參閱使用 AWS Glue和 Lake Formation 管理跨帳戶許可。
- 如果您目前正在使用 AWS Glue Data Catalog 資源政策來共用資源，而且想要使用跨帳戶版本設定第 3 版授予跨帳戶許可，則必須使用 `glue:PutResourcePolicy` API 操作在資料目錄設定中新增`glue:ShareResource`許可，如 使用 AWS Glue和 Lake Formation 管理跨帳戶許可一節所示。如果您的帳戶未使用 AWS Glue Data Catalog 資源政策 （ 第 1 版和第 2 版使用`glue:PutResourcePolicy`許可 ） 授予跨帳戶存取權，則不需要此政策。

```
{
      "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"
      ]
    }
```

- 如果您的帳戶已使用 AWS Glue Data Catalog 資源政策進行跨帳戶共用，且您目前使用具名資源方法或 LF-TBAC 搭配跨帳戶設定第 3 版來共用資源，則當您叫用 `glue:PutResourcePolicy` API 操作`'true'`時 AWS RAM ，必須將 `EnableHybrid`引數設定為 。如需詳細資訊，請參閱使用 AWS Glue和 Lake Formation 管理跨帳戶許可。

存取共用資源的每個帳戶中所需的設定

- 如果您要與 共用資源 AWS 帳戶，則取用者帳戶中至少有一個使用者必須是資料湖管理員，才能檢視共用資源。如需如何建立資料湖管理員的資訊，請參閱 建立資料湖管理員。

  資料湖管理員可以將共用資源的 Lake Formation 許可授予帳戶中的其他主體。在資料湖管理員授予資源許可之前，其他主體無法存取共用資源。
- Athena 和 Redshift Spectrum 等整合服務需要資源連結，才能在查詢中包含共用資源。委託人需要在其 Data Catalog 中建立資源連結，以從另一個 共用資源 AWS 帳戶。如需資源連結的詳細資訊，請參閱 資源連結在 Lake Formation 中如何運作。
- 當資源直接與 IAM 主體共用時，若要使用 Athena 查詢資料表，主體需要建立資源連結。若要建立資源連結，委託人需要 Lake Formation `CREATE_TABLE`或 `CREATE_DATABASE` 許可，以及 `glue:CreateTable`或 IAM `glue:CreateDatabase` 許可。

  如果生產者帳戶在相同資料庫中與相同或另一個委託人共用不同的資料表，則該委託人可以立即查詢資料表。

> **Note**
>
> 對於資料湖管理員和資料湖管理員已授予許可的主體，共用資源會顯示在資料目錄中，就像是本機（擁有）資源一樣。擷取、轉換和載入 (ETL) 任務可以存取共用資源的基礎資料。
> 對於共用資源，Lake Formation 主控台上的資料表和資料庫頁面會顯示擁有者的帳戶 ID。
> 存取共用資源的基礎資料時，CloudTrail 日誌事件會在共用資源收件人的帳戶和資源擁有者帳戶中產生。CloudTrail 事件可以包含存取資料的主體 ARN，但前提是收件人帳戶選擇在日誌中包含主體 ARN。如需詳細資訊，請參閱跨帳戶 CloudTrail 記錄。

更新跨帳戶資料共用版本設定

會不時 AWS Lake Formation 更新跨帳戶資料共用設定，以區分對 AWS RAM 用量所做的變更，並支援對跨帳戶資料共用功能所做的更新。當 Lake Formation 執行此操作時，它會建立新的跨帳戶版本設定版本。

跨帳戶版本設定的主要差異

如需跨帳戶資料共用在不同跨帳戶版本設定下運作方式的詳細資訊，請參閱下列各節。

> **Note**
>
> 若要與其他帳戶共用資料，授予者必須具有`AWSLakeFormationCrossAccountManager`受管 IAM 政策許可。這是所有版本的先決條件。
> 更新跨帳戶版本設定不會影響收件人對共用資源所擁有的許可。這適用於從第 1 版更新到第 2 版、將第 2 版更新到第 3 版，以及將第 1 版更新到第 3 版。更新版本時，請參閱下列考量事項。

第 1 版

具名資源方法：將每個跨帳戶 Lake Formation 許可授予映射至一個 AWS RAM 資源共享。使用者（授權方角色或委託人）不需要額外的許可。

LF-TBAC 方法：跨帳戶 Lake Formation 許可授予不會用來 AWS RAM 共用資料。使用者必須擁有 `glue:PutResourcePolicy` 許可。

更新版本的好處：初始版本 - 不適用。

更新版本時的考量事項：初始版本 - 不適用

2 版

具名資源方法：透過映射具有一個 AWS RAM 資源共享的多個跨帳戶許可授予來最佳化 AWS RAM 資源共享的數量。使用者不需要額外的許可。

LF-TBAC 方法：跨帳戶 Lake Formation 許可授予不會用來 AWS RAM 共用資料。使用者必須擁有 `glue:PutResourcePolicy` 許可。

更新版本的優勢：透過 AWS RAM 容量的最佳使用率進行可擴展的跨帳戶設定。

更新版本時的考量：想要授予跨帳戶 Lake Formation 許可的使用者，必須在 AWSLakeFormationCrossAccountManager AWS 受管政策中擁有 許可。否則，您需要有 ram:AssociateResourceShare和 ram:DisassociateResourceShare許可，才能成功與其他 帳戶共用資源。

第 3 版

具名資源方法：透過映射具有一個 AWS RAM 資源共享的多個跨帳戶許可授予，來最佳化 AWS RAM 資源共享的數量。使用者不需要額外的許可。

LF-TBAC 方法：Lake Formation 使用 AWS RAM 進行跨帳戶授予。使用者必須將 glue：ShareResource 陳述式新增至glue:PutResourcePolicy許可。收件人必須接受來自 的資源共用邀請 AWS RAM。

更新版本的優勢： 支援下列功能：

- 允許與外部帳戶中的 IAM 主體明確共用資源。

  如需詳細資訊，請參閱授予 Data Catalog 資源的許可。
- 使用 LF-TBAC 方法對 Organizations 或組織單位 (OUs) 啟用跨帳戶共用。
- 移除維護跨帳戶授與額外 AWS Glue 政策的額外負荷。

更新版本時的考量事項：當您使用 LF-TBAC 方法共用資源時，如果授予者使用低於第 3 版的版本，且收件人使用第 3 版或更新版本，則授予者會收到下列錯誤訊息：「無效的跨帳戶授予請求。消費者帳戶可選擇加入跨帳戶版本：v3。請將 CrossAccountVersion 中的 更新DataLakeSetting為最低版本 v3 （服務：AmazonDataCatalog；狀態碼：400；錯誤碼：InvalidInputException)"。不過，如果授予者使用第 3 版，且收件人使用第 1 版或第 2 版，則使用 LF 標籤的跨帳戶授予會順利通過。

使用具名資源方法進行的跨帳戶授予在不同版本之間相容。即使授予者帳戶使用較舊版本 （第 1 版或第 2 版） 且收件人帳戶使用較新版本 （第 3 版或更高版本），跨帳戶存取功能也能順暢運作，而不會發生任何相容性問題或錯誤。

若要直接與另一個帳戶中的 IAM 主體共用資源，只有授予者需要使用第 3 版。

使用 LF-TBAC 方法進行的跨帳戶授予需要使用者在帳戶中擁有 AWS Glue Data Catalog 資源政策。當您更新到第 3 版時，LF-TBAC 授予會使用 AWS RAM。若要允許以 AWS RAM 為基礎的跨帳戶授予成功，您必須將 glue:ShareResource陳述式新增至現有的 Data Catalog 資源政策，如使用 AWS Glue和 Lake Formation 管理跨帳戶許可一節所示。

第 4 版

授予者需要第 4 版或更新版本，才能在混合存取模式中共用 Data Catalog 資源，或在聯合目錄中共用物件。

最佳化 AWS RAM 資源共用

跨帳戶授與的新版本（第 2 版及更高版本）會最佳化利用 AWS RAM 容量來最大化跨帳戶用量。當您與外部 AWS 帳戶 或 IAM 主體共用資源時，Lake Formation 可能會建立新的資源共用，或將資源與現有共用建立關聯。透過與現有共享建立關聯，Lake Formation 可減少消費者需要接受的資源共享邀請數量。

透過 TBAC 啟用 AWS RAM 共用或直接與委託人共用資源

若要直接與另一個帳戶中的 IAM 主體共用資源，或啟用對 Organizations 或組織單位的 TBAC 跨帳戶共用，您需要將跨帳戶版本設定更新為第 3 版。如需 AWS RAM 資源限制的詳細資訊，請參閱 跨帳戶資料共用最佳實務和考量事項。

更新跨帳戶版本設定所需的許可

如果跨帳戶許可授予者具有AWSLakeFormationCrossAccountManager受管 IAM 政策許可，則跨帳戶許可授予者角色或委託人不需要額外的許可設定。不過，如果跨帳戶授予者未使用 受管政策，則授予者角色或委託人應具有下列 IAM 許可，才能讓跨帳戶授予的新版本成功。

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "VisualEditor1",
      "Effect": "Allow",
      "Action": [
         "ram:AssociateResourceShare",
         "ram:DisassociateResourceShare",
         "ram:GetResourceShares"
       ],
     "Resource": "*",
     "Condition": {
       "StringLike": {
         "ram:ResourceShareName": "LakeFormation*"
        }
```

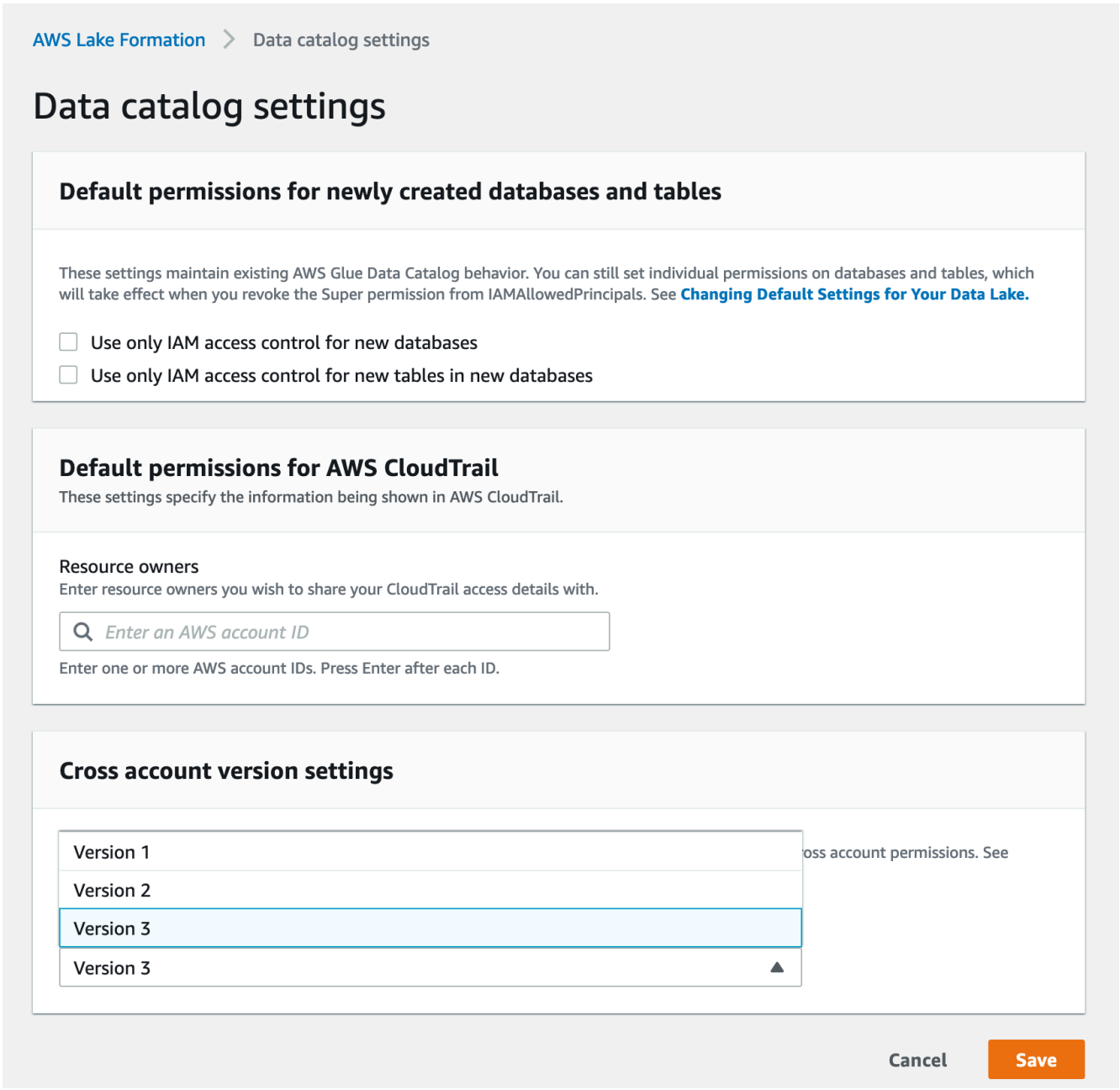
```
      }
    }
  ]
}
```

啟用新版本

請依照下列步驟，透過 AWS Lake Formation 主控台或 更新跨帳戶版本設定 AWS CLI。

Console

1. 在資料目錄設定頁面的跨帳戶版本設定下，選擇第 2 版、第 3 版或第 4 版。如果您選取第 1 版，Lake Formation 將使用預設資源共用模式。



2. 選擇儲存。

AWS Command Line Interface (AWS CLI)

使用 `put-data-lake-settings` AWS CLI 命令來設定 `CROSS_ACCOUNT_VERSION` 參數。接受的值為 1、2、3 和 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**
>
> 選擇第 2 版或第 3 版後，所有新的具名資源授予都會經過新的跨帳戶授予模式。若要最佳地使用現有跨帳戶共用的 AWS RAM 容量，建議您撤銷使用較舊版本進行的授予，並在新模式中重新授予。

從外部帳戶跨 AWS 帳戶 或 IAM 主體共用 Data Catalog 資料表和資料庫

本節包含如何將 Data Catalog 資源的跨帳戶許可授予外部 AWS 帳戶、IAM 主體、 AWS 組織或組織單位的指示。授予操作會自動共用這些資源。

主題

- 使用標籤型存取控制進行資料共用
- 使用具名資源方法的跨帳戶資料共用

使用標籤型存取控制進行資料共用

AWS Lake Formation 標籤型存取控制 (LF-TBAC) 是一種授權策略，可根據屬性定義許可。下列步驟說明如何使用 LF 標籤授予跨帳戶許可。

在生產者/生產者帳戶上設定必要項目

1. 新增 LF 標籤。

    1. 以資料湖管理員或 LF 標籤建立者身分登入 Lake Formation 主控台。

2. 在左側導覽列中，選擇許可，以及 LF 標籤和許可。
3. 選擇新增 LF 標籤。

   如需建立 LF 標籤的詳細說明，請參閱 建立 LF 標籤。

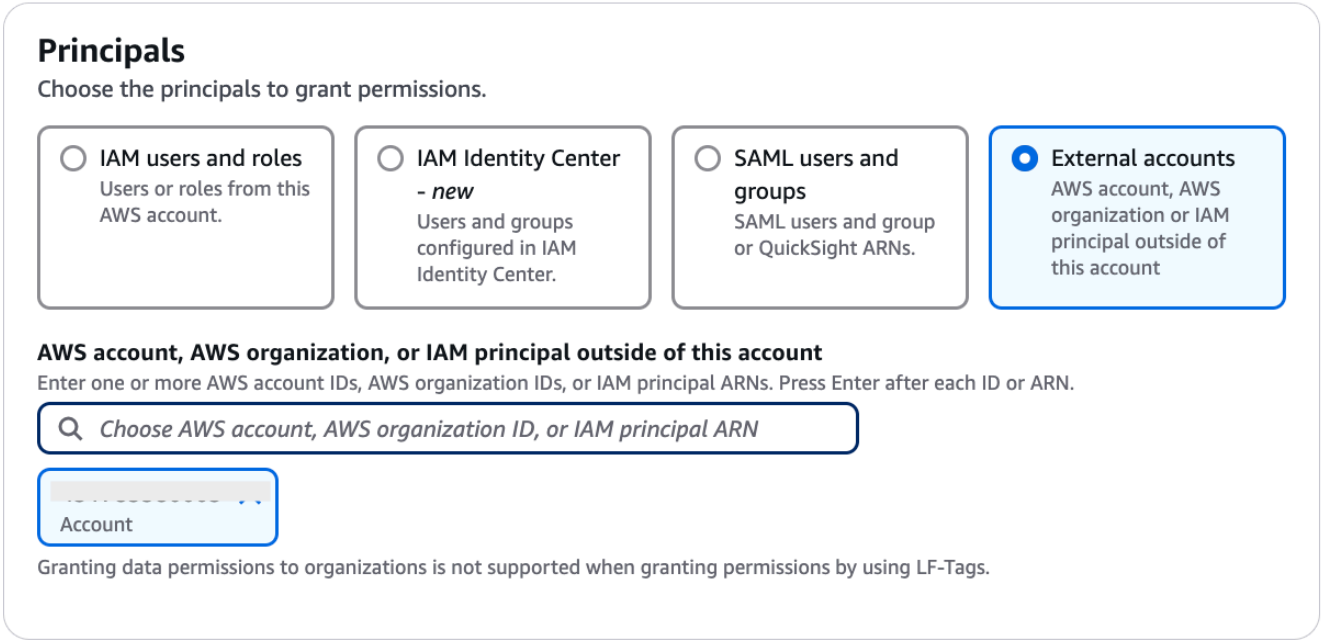
2. 授予 描述和/或關聯許可 LF 標籤鍵/值對至您帳戶或外部帳戶中的 IAM 主體。

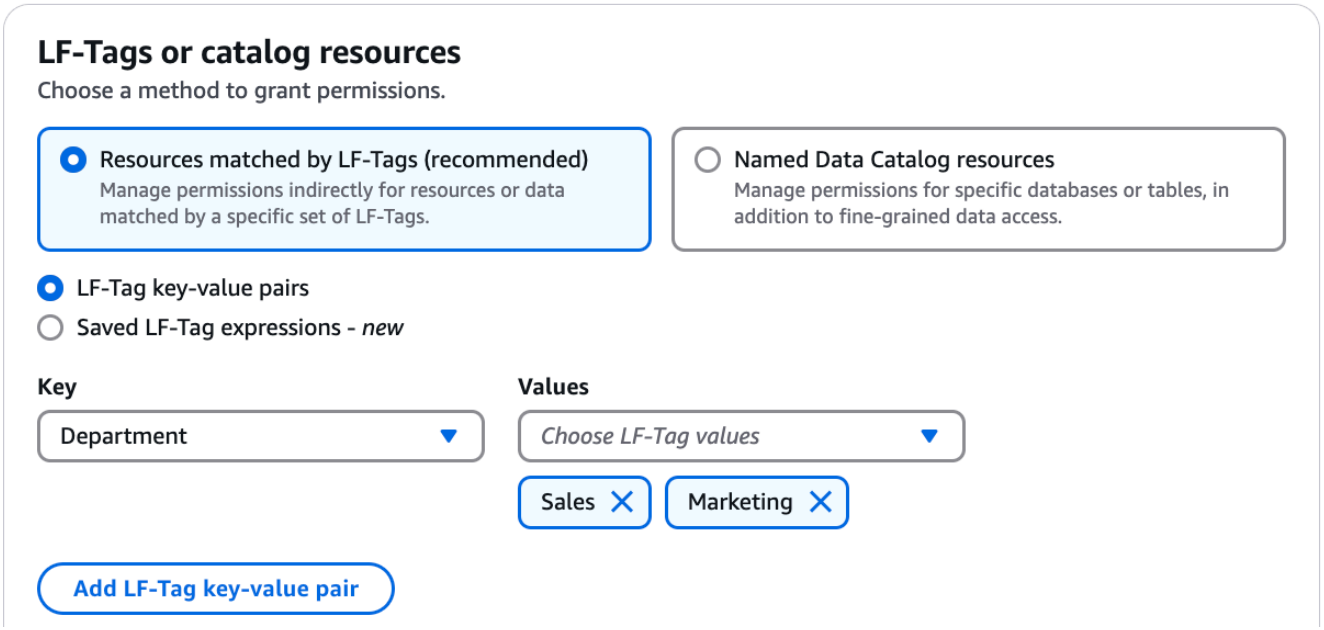
   授予 LF-Tag 鍵/值對的許可可讓主體檢視 LF-Tag，並將其指派給 Data Catalog 資源 （資料庫、資料表和資料欄）。

3. 接著，資料湖管理員或具有關聯許可的 IAM 主體可以將 LF 標籤指派給資料庫、資料表或資料欄。如需詳細資訊，請參閱將 LF 標籤指派給 Data Catalog 資源。

4. 接著，使用 LF-Tag 表達式將資料許可授予外部帳戶。這可讓許可的承授者或接收者存取以相同（多個）金鑰和 （多個） 值標記的 Data Catalog 資源。

   1. 在導覽窗格中，選擇許可和資料許可。
   2. 選擇 Grant (授予)。
   3. 在授予許可頁面上，針對委託人，選擇外部帳戶，然後輸入委託人的承授者 AWS 帳戶 ID 或 IAM 角色，或針對委託人 （委託人 ARN) 輸入 Amazon Resource Name (ARN)，以直接跨帳戶授予外部委託人。輸入帳戶 ID 後，您需要按 Enter。





4. 針對 LF 標籤或目錄資源，選擇符合 LF 標籤的資源 （ 建議 ）。

   a. 選擇選項 LF-Tag 鍵/值對或已儲存的 LF-Tag 表達式 。

   b. 如果您選擇 LF-Tag 鍵/值對，請輸入與要與承授者帳戶共用之 Data Catalog 資源相關聯的 LF-Tag 鍵和值 (LF-Tag)。

      被授予者在 LF-Tag 表達式中獲指派相符 LF-Tag 之 Data Catalog 資源的許可。如果 LF-Tag 表達式指定每個標籤索引鍵的多個值，則任何一個標籤值都可以相符。

5. 選擇資料庫層級或資料表層級許可，在符合 LF-Tag 表達式的資源上授予 。

> **Important**
>
> 由於資料湖管理員必須將共用資源的許可授予承授者帳戶中的主體，因此您必須一律使用授予選項授予跨帳戶許可。

如需詳細資訊，請參閱使用主控台授予 LF-Tag 許可。

> **Note**
>
> 收到直接跨帳戶授予的委託人將沒有可授予許可選項。

在接收/承授者帳戶上設定必要

1. 以消費者帳戶的資料湖管理員身分登入 Lake Formation 主控台。
2. 接著，接收消費者帳戶中的資源共享。
   1. 開啟 AWS RAM 主控台。
   2. 在導覽窗格中，選擇與我共用下的資源共用。
   3. 選取資源共用，選擇接受資源共用。
3. 當您與其他帳戶共用資源時，資源仍屬於生產者帳戶，且在 Athena 主控台中看不到。若要在 Athena 主控台中顯示資源，您需要建立指向共用資源的資源連結。如需建立資源連結的說明，請參閱 建立共用 Data Catalog 資料表的資源連結 和 建立共用 Data Catalog 資料庫的資源連結
   1. 在資料目錄下選擇資料庫或資料表。
   2. 在資料庫/資料表頁面上，選擇建立、資源連結 。
   3. 為資料庫資源連結輸入下列資訊：
      - 資源連結名稱 – 資源連結的唯一名稱。
      - 目的地目錄 – 您要建立資源連結的目錄。
      - 共用資料庫區域 – 如果您要在不同區域中建立資源連結，則與您共用的資料庫區域。
      - 共用資料庫 – 選擇共用資料庫。
      - 共用資料庫的目錄 ID – 輸入共用資料庫的目錄 ID。
   4. 選擇建立。您可以在資料庫清單中看到新建立的資源連結。

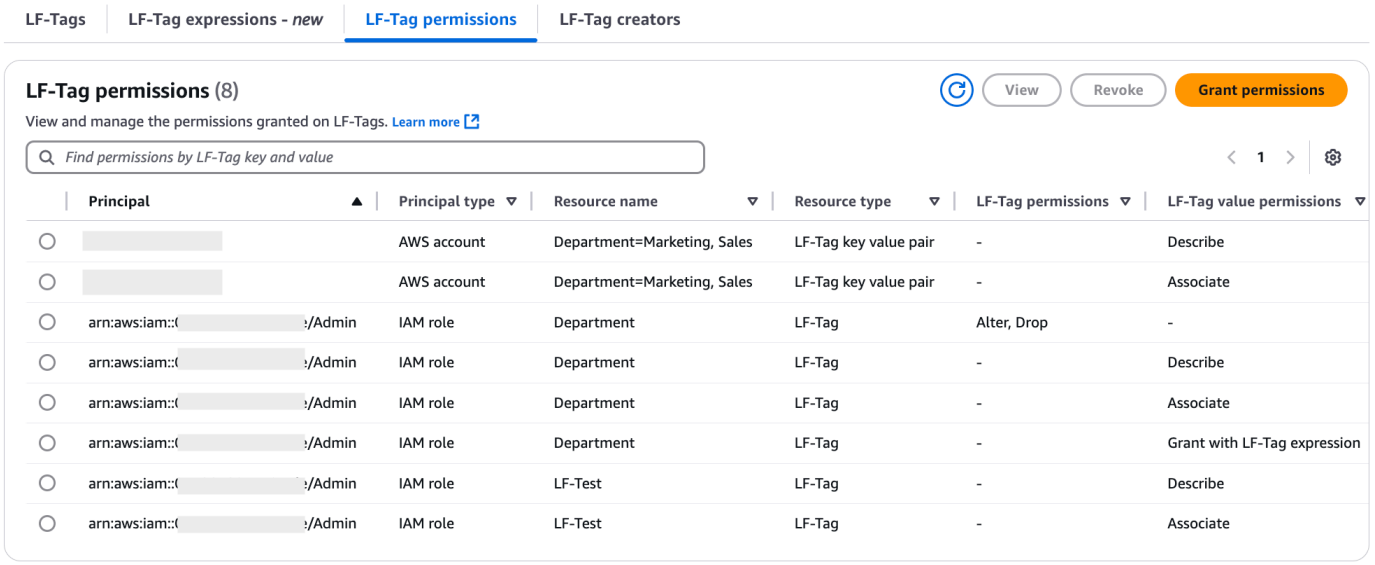
同樣地，您可以建立共用資料表的資源連結。

4. 現在授予您要共用資源之 IAM 主體的資源連結描述許可。

   1. 在資料庫/資料表頁面上，選取資源連結，然後在動作功能表中選擇授予。
   2. 在授予許可區段中，選取 IAM 使用者和角色。
   3. 選擇您要授予資源連結存取權的 IAM 角色。
   4. 在資源連結許可區段中，選取描述。
   5. 選擇 Grant (授予)。

5. 接著，將 LF-Tag 鍵值許可授予取用者帳戶中的主體。

   您應該能夠在 Lake Formation 主控台的許可、LF 標籤和許可下，找到消費者帳戶中與您共用的 LF 標籤。您可以將從授予者共用的標籤與從授予者帳戶共用的資源建立關聯，其中包括：資料庫、資料表和資料欄。您可以進一步將資源的許可授予其他委託人。



   1. 在導覽窗格的許可、資料許可下，選擇授予。
   2. 在授予許可頁面上，選擇 IAM 使用者和角色。
   3. 接著，選擇您帳戶中的 IAM 使用者和角色，以授予共用資料庫/資料表的存取權。
   4. 接著，針對 LF 標籤或目錄資源，選擇符合 LF 標籤的資源。
   5. 接著，選擇與您共用之 LF 標籤的金鑰和值。
   6. 接著，選擇您要授予 IAM 使用者和角色的資料庫和資料表許可。您也可以選擇可授予許可，讓 IAM 使用者和角色將許可授予其他使用者/角色。

7. 選擇 Grant (授予)。
8. 您可以在 Lake Formation 主控台的資料許可下檢視許可授予。

使用具名資源方法的跨帳戶資料共用

您可以直接將許可授予另一個 AWS 帳戶中的委託人，或授予外部 AWS 帳戶 或 AWS Organizations。將 Lake Formation 許可授予 Organizations 或組織單位等同於將許可授予該組織或組織單位 AWS 帳戶 中的每個 。

當您將許可授予外部帳戶或組織時，您必須包含可授予許可選項。只有外部帳戶中的資料湖管理員可以存取共用資源，直到管理員將共用資源的許可授予外部帳戶中的其他主體為止。

> **Note**
>
> 從外部帳戶將許可直接授予 IAM 主體時，不支援可授予的許可選項。

遵循 中的指示使用具名資源方法授予資料庫許可，使用具名資源方法授予跨帳戶許可。

授予與您帳戶共用之資料庫或資料表的許可

與您的帳戶 AWS 共用屬於另一個 AWS 帳戶的 Data Catalog 資源之後，身為資料湖管理員，您可以將共用資源的許可授予帳戶中的其他主體。不過，您無法將資源的許可授予其他 AWS 帳戶或組織。

您可以使用 AWS Lake Formation 主控台、 API 或 AWS Command Line Interface (AWS CLI) 來授予許可。

授予共用資料庫的許可 （命名資源方法、主控台）

- 請遵循中的說明進行使用具名資源方法授予資料庫許可 在 LF 標籤或目錄資源下的資料庫清單中，請確定您在外部帳戶中選取資料庫，而不是資料庫的資源連結。

  如果您在資料庫清單中看不到資料庫，請確定您已接受資料庫的 AWS Resource Access Manager (AWS RAM) 資源共享邀請。如需詳細資訊，請參閱從 接受資源共享邀請 AWS RAM。

  此外，對於 CREATE_TABLE和 ALTER許可，請遵循 中的指示授予資料位置許可 （相同帳戶），並務必在註冊帳戶位置欄位中輸入擁有的帳戶 ID。

授予共用資料表的許可（命名資源方法、主控台）

- 請遵循中的說明進行使用具名資源方法授予資料表許可 在 LF 標籤或目錄資源下的資料庫清單中，請確定您在外部帳戶中選取資料庫，而不是資料庫的資源連結。

  如果您在資料表清單中看不到資料表，請確定您已接受 AWS RAM 資料表的資源共享邀請。如需詳細資訊，請參閱從 接受資源共享邀請 AWS RAM。

  此外，對於 ALTER許可，請遵循 中的指示授予資料位置許可（相同帳戶），並務必在註冊帳戶位置欄位中輸入擁有的帳戶 ID。

若要授予共用資源的許可 (LF-TBAC 方法、主控台）

- 請遵循中的說明進行授予 Data Catalog 許可 在 LF-Tags 或目錄資源區段中，授予外部帳戶授予您帳戶的確切 LF-Tag 表達式，或該表達式的子集。

  例如，如果外部帳戶使用授予選項將 LF-Tag 運算式授予module=customers AND environment=production您的帳戶，則身為資料湖管理員，您可以將該相同運算式或 module=customers environment=production 授予您帳戶中的委託人。您只能授予透過 LF-Tag 表達式授予資源的相同或一部分 Lake Formation 許可（例如 ALTER、 SELECT等）。

授予共用資料表的許可（命名資源方法）AWS CLI

- 輸入與以下相似的命令。在此範例中：
  - AWS 您的帳戶 ID 為 1111-2222-3333。
  - 擁有資料表並將其授予您帳戶的 帳戶為 1234-5678-9012。
  - 正在共用資料表上將 SELECT許可授予pageviews使用者 datalake_user1。該使用者是您帳戶中的委託人。
  - pageviews 資料表位於 analytics 資料庫中，其為帳戶 1234-5678-9012 所擁有。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "SELECT" --resource '{ "Table": {"CatalogId":"123456789012",
 "DatabaseName":"analytics", "Name":"pageviews"}}'
```

  請注意，必須在 resource引數的 CatalogId 屬性中指定擁有帳戶。

授予資源連結許可

請依照下列步驟，將一或多個資源連結的 AWS Lake Formation 許可授予您 AWS 帳戶中的委託人。

建立資源連結之後，只有您可以檢視和存取它。（這假設未針對資料庫啟用此資料庫中的新資料表僅使用 IAM 存取控制。）若要允許帳戶中的其他主體存取資源連結，請授予至少 DESCRIBE許可。

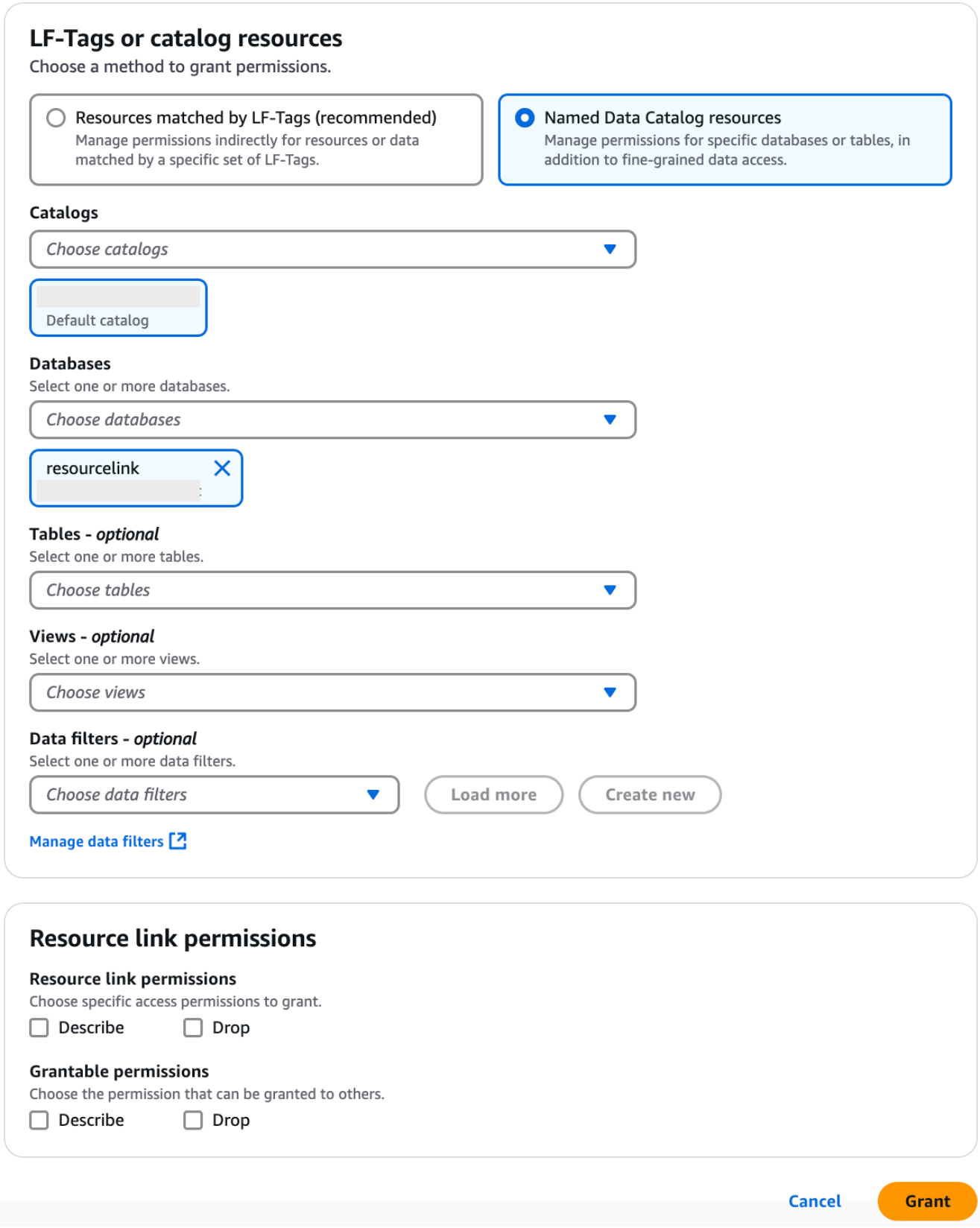
> **Important**
>
> 授予資源連結的許可不會授予目標（連結）資料庫或資料表的許可。您必須分別授予目標的許可。

您可以使用 Lake Formation 主控台、API 或 AWS Command Line Interface () 來授予許可AWS CLI。

console

使用 Lake Formation 主控台授予資源連結許可

1. 執行以下任意一項：
   - 對於資料庫資源連結，請遵循 中的步驟使用具名資源方法授予資料庫許可。 執行下列動作：
     1. 從 Data Catalog 資料庫下的資料庫清單中選取資源連結。
     2. 選擇授予以開啟授予許可頁面。
     3. 指定要授予許可的委託人。
     4. 目錄和資料庫欄位會填入。
   - 對於資料表資源連結，請依照 中的步驟使用具名資源方法授予資料表許可執行下列動作：
     1. 從 Data Catalog， Tables 下的資料表清單中選取資源連結。
     2. 開啟授予許可頁面。
     3. 指定主體。
     4. 目錄、資料庫、資料表欄位會填入。
     5. 指定主體。
2. 在許可下，選取要授予的許可。或者，選取可授予的許可。


LF-Tags or catalog resources
Choose a method to grant permissions.
Resources matched by LF-Tags (recommended)
Manage permissions indirectly for resources or data matched by a specific set of LF-Tags.
Named Data Catalog resources
Manage permissions for specific databases or tables, in addition to fine-grained data access.
Catalogs
Choose catalogs
Default catalog
Databases
Select one or more databases.
Choose databases
resourcelink
Tables - optional
Select one or more tables.
Choose tables
Views - optional
Select one or more views.
Choose views
Data filters - optional
Select one or more data filters.
Choose data filters
Load more
Create new
Manage data filters
Resource link permissions
Resource link permissions
Choose specific access permissions to grant.
Describe
Drop
Grantable permissions
Choose the permission that can be granted to others.
Describe
Drop
Cancel
Grant

3. 選擇 Grant (授予)。

AWS CLI

使用 授予資源連結許可 AWS CLI

- 執行 `grant-permissions`命令，指定資源連結做為資源。

  Example

  此範例`DESCRIBE`會授予`issues` AWS 帳戶 1111-2222-3333 `incidents-link` 中資料庫中資料表資源連結`datalake_user1`上的使用者。

```
aws lakeformation grant-permissions --principal
 DataLakePrincipalIdentifier=arn:aws:iam::111122223333:user/datalake_user1
 --permissions "DESCRIBE" --resource '{ "Table": {"DatabaseName":"issues",
 "Name":"incidents-link"}}'
```

另請參閱:

- 建立資源連結
- Lake Formation 許可參考

存取共用資料表的基礎資料

假設 AWS 帳戶 A 與帳戶 B 共用 Data Catalog 資料表，例如，`SELECT`透過將資料表上的授予選項授予帳戶 B。若要讓帳戶 B 中的主體能夠讀取共用資料表的基礎資料，必須符合下列條件：

- 帳戶 B 中的資料湖管理員必須接受共用。（如果帳戶 A 和 B 位於相同組織中，或使用 Lake Formation 標籤型存取控制方法進行授予，則不需要這麼做。）
- 資料湖管理員必須向委託人重新授予共用資料表上帳戶 A 的 Lake Formation `SELECT`許可。
- 委託人必須在資料表、包含它的資料庫以及帳戶 A Data Catalog 上具有下列 IAM 許可。

Note

在下列 IAM 政策中：

- 以 AWS 帳戶 A 的帳戶 ID 取代 *<account-id-A>*。
- 將 *<region>* 取代為有效的區域。

- 將 *<database>* 取代為帳戶 A 中包含共用資料表的資料庫名稱。
- 將 *<table>* 取代為共用資料表的名稱。

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

另請參閱:

- 從 接受資源共享邀請 AWS RAM

跨帳戶 CloudTrail 記錄

Lake Formation 提供對資料湖中資料的所有跨帳戶存取的集中式稽核線索。當收件人 AWS 帳戶存取共用資料表中的資料時，Lake Formation 會將 CloudTrail 事件複製到擁有帳戶的 CloudTrail 日誌。複製的事件包括透過整合服務查詢資料，例如 Amazon Athena 和 Amazon Redshift Spectrum，以及依 AWS Glue任務存取資料。

Data Catalog 資源上跨帳戶操作的 CloudTrail 事件也會類似地複製。

身為資源擁有者，如果您在 Amazon S3 中啟用物件層級記錄，您可以執行將 S3 CloudTrail 事件與 Lake Formation CloudTrail 事件聯結的查詢，以判斷已存取 S3 儲存貯體的帳戶。

主題

- 在跨帳戶 CloudTrail 日誌中包含主體身分
- 查詢 Amazon S3 跨帳戶存取的 CloudTrail 日誌

在跨帳戶 CloudTrail 日誌中包含主體身分

根據預設，跨帳戶 CloudTrail 事件會新增至共用資源收件人的日誌，並複製到資源擁有者的日誌，只包含外部帳戶主體的 AWS 主體 ID，而不是主體的人類可讀取 Amazon Resource Name (ARN)（主體 ARN)。在信任的界限內共用資源時，例如同一組織或團隊內，您可以選擇在 CloudTrail 事件中包含委託人 ARN。然後，資源擁有者帳戶可以追蹤收件人帳戶中存取其擁有資源的主體。

> **Important**
>
> 身為共用資源收件人，若要在您自己的 CloudTrail 日誌中查看事件中的委託人 ARN，您必須選擇與擁有者帳戶共用委託人 ARN。
>
> 如果資料存取是透過資源連結進行，則會在共用資源收件人帳戶中記錄兩個事件：一個用於資源連結存取，另一個用於目標資源存取。資源連結存取的事件包含委託人 ARN。目標資源存取的事件不包含沒有選擇加入的委託人 ARN。資源連結存取事件不會複製到擁有者帳戶。

以下是預設跨帳戶 CloudTrail 事件（未選擇加入）的摘錄。執行資料存取的帳戶為 1111-2222-3333。這是顯示在呼叫帳戶和資源擁有者帳戶中的日誌。Lake Formation 會在跨帳戶案例中的兩個帳戶中填入日誌。

```
{
    "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"
    },
...
}
```

身為共用資源取用者，當您選擇加入委託人 ARN 時，摘錄會變成以下內容。`lakeFormationPrincipal` 欄位代表透過 Amazon Athena、Amazon Redshift Spectrum 或 AWS Glue任務執行查詢的最終角色或使用者。

```
{
    "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"
    },
...
}
```

選擇在跨帳戶 CloudTrail 日誌中包含主體 ARNs

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

以Administrator使用者身分登入，或使用 IAM Administrator Access 政策的使用者登入。

2. 在導覽窗格中，選擇設定。
3. 在資料目錄設定頁面上，於資源擁有者的預設許可 AWS CloudTrail區段中，輸入一或多個 AWS 資源擁有者帳戶 IDs。

   在每個帳戶 ID 之後按 Enter。
4. 選擇儲存。

   現在，跨帳戶 CloudTrail 事件存放在共用資源收件人和資源擁有者的日誌中，包含主體 ARN。

查詢 Amazon S3 跨帳戶存取的 CloudTrail 日誌

身為共用資源擁有者，您可以查詢 S3 CloudTrail 日誌，以判斷已存取 Amazon S3 儲存貯體的帳戶（前提是您在 Amazon S3 中啟用物件層級記錄）。這僅適用於您向 Lake Formation 註冊的 S3 位置。如果共用資源消費者選擇在 Lake Formation CloudTrail 日誌中包含主體 ARNs，您可以判斷存取儲存貯體的角色或使用者。

使用 執行查詢時 Amazon Athena，您可以在工作階段名稱屬性上加入 Lake Formation CloudTrail 事件和 S3 CloudTrail 事件。查詢也可以篩選 上的 Lake Formation 事件eventName="GetDataAccess"，以及 eventName="Get Object"或 上的 S3 事件eventName="Put Object"。

以下是 Lake Formation 跨帳戶 CloudTrail 事件的摘錄，其中已存取已註冊 S3 位置中的資料。

```
{
  "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"
   }
}
```

lakeFormationRoleSessionName 金鑰值 AWSLF-00-GL-111122223333-B8JSAjo5QA可與 S3 CloudTrail 事件的principalId金鑰中的工作階段名稱聯結。以下是 S3 CloudTrail 事件的摘錄。它會顯示工作階段名稱的位置。

```
{
   "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"
     },
   ..............
   ..............
}
```

工作階段名稱的格式如下：

```
AWSLF-<version-number>-<query-engine-code>-<account-id->-<suffix>
```

version-number

此格式的版本，目前為 00。如果工作階段名稱格式變更，下一個版本將為 01。

query-engine-code

指出存取資料的實體。目前的值為：

GL	AWS Glue ETL 任務
AT	Athena

RE	Amazon Redshift Spectrum

account-id

從 Lake Formation 請求憑證 AWS 的帳戶 ID。

suffix

隨機產生的字串。

使用 AWS Glue和 Lake Formation 管理跨帳戶許可

您可以使用 AWS Glue或 授予 Data Catalog 資源和基礎資料的跨帳戶存取權 AWS Lake Formation。

在 中AWS Glue，您可以透過建立或更新 Data Catalog 資源政策來授予跨帳戶許可。在 Lake Formation 中，您可以使用 Lake Formation 許可模型和 `Grant Permissions` API 操作來授予跨帳戶GRANT/REVOKE許可。

> **Tip**
>
> 我們建議僅依賴 Lake Formation 許可來保護您的資料湖。

您可以使用 Lake Formation 主控台或 AWS Resource Access Manager (AWS RAM) 主控台來檢視 Lake Formation 跨帳戶授予。不過，這些主控台頁面不會顯示 AWS Glue Data Catalog 資源政策授予的跨帳戶許可。同樣地，您可以使用AWS Glue主控台的設定頁面檢視 Data Catalog 資源政策中的跨帳戶授予，但該頁面不會顯示使用 Lake Formation 授予的跨帳戶許可。

為了確保您在檢視和管理跨帳戶許可時不會錯過任何授予，Lake Formation 和 AWS Glue會要求您執行下列動作，以表示您知道並允許 Lake Formation 和 進行跨帳戶授予AWS Glue。

使用 AWS Glue Data Catalog 資源政策授予跨帳戶許可時

如果您的帳戶（匯款人帳戶或生產者帳戶）未授予使用 AWS RAM 來共用資源的跨帳戶授權，您可以像往常一樣在 中儲存 Data Catalog 資源政策AWS Glue。不過，如果已進行涉及 AWS RAM 資源共享的授予，您必須執行下列其中一項操作，以確保成功儲存資源政策：

- 當您在AWS Glue主控台的設定頁面上儲存資源政策時，主控台會發出提醒，指出政策中的許可除了使用 Lake Formation 主控台授予的任何許可之外。您必須選擇繼續以儲存政策。

- 使用 `glue:PutResourcePolicy` API 操作儲存資源政策時，您必須將 `EnableHybrid` 欄位設定為 'TRUE'（類型 = 字串）。下列程式碼範例示範如何在 Python 中執行此操作。

```
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')
```

如需詳細資訊，請參閱《 AWS Glue 開發人員指南》中的 PutResourcePolicy 動作 (Python：put_resource_policy)。

使用 Lake Formation 命名的資源方法授予跨帳戶許可時

如果您的帳戶（生產者帳戶）中沒有 Data Catalog 資源政策，您照常進行 Lake Formation 跨帳戶授予。不過，如果 Data Catalog 資源政策存在，您必須將下列陳述式新增至其中，以允許跨帳戶授予在使用具名資源方法進行時成功。將 *<region>* 取代為有效的區域名稱，並將 *<account-id>* 取代為 AWS 您的帳戶 ID（生產者帳戶 ID)。

```
{
  "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"
  ]
}
```

如果沒有此額外陳述式，Lake Formation 授予會成功，但會遭到封鎖 AWS RAM，而且收件人帳戶無法存取授予的資源。

Important

使用 Lake Formation 標籤型存取控制 (LF-TBAC) 方法進行跨帳戶授予時，您必須擁有至少具有 中指定許可的 Data Catalog 資源政策先決條件。

另請參閱:

- 中繼資料存取控制（討論具名資源方法與 Lake Formation 標籤型存取控制 (LF-TBAC) 方法）。
- 檢視共用的 Data Catalog 資料表和資料庫
- 《AWS Glue 開發人員指南》中的在AWS Glue主控台上使用資料目錄設定
- 《AWS Glue 開發人員指南》中的授予跨帳戶存取權（適用於範例 Data Catalog 資源政策）

使用 GetResourceShares API 操作檢視所有跨帳戶授予

如果您的企業同時使用 AWS Glue Data Catalog 資源政策和 Lake Formation 授予來授予跨帳戶許可，在同一個位置檢視所有跨帳戶授予的唯一方法是使用 `glue:GetResourceShares` API 操作。

當您使用具名資源方法跨帳戶授予 Lake Formation 許可時， AWS Resource Access Manager (AWS RAM) 會建立 AWS Identity and Access Management (IAM) 資源政策，並將其存放在 AWS 您的帳戶中。政策會授予存取 資源所需的許可。 會為每個跨帳戶授予 AWS RAM 建立個別的資源政策。您可以使用 `glue:GetResourceShares` API 操作來檢視所有這些政策。

> **Note**
>
> 此操作也會傳回 Data Catalog 資源政策。不過，如果您在 Data Catalog 設定中啟用中繼資料加密，而且您沒有 AWS KMS 金鑰的許可，則操作不會傳回 Data Catalog 資源政策。

檢視所有跨帳戶授予

- 輸入下列 AWS CLI 命令。

```
aws glue get-resource-policies
```

以下是範例資源政策，當您db1將t資料庫中資料表的許可授予 AWS 帳戶 1111-2222-3333 時， AWS RAM 會建立和存放該政策。

```
{
  "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"
      ]
     }
   ]
}
```

> 另請參閱：
>
> - 《AWS Glue 開發人員指南》中的 GetResourceShares 動作 (Python：get_resource_policies)

存取和檢視共用的 Data Catalog 資料表和資料庫

對於資料湖管理員和已授予許可的主體，與 AWS 您的帳戶共用的資源會顯示在 Data Catalog 中，就像是帳戶中的資源一樣。主控台會顯示擁有資源的帳戶。

您可以使用 Lake Formation 主控台來檢視與您的帳戶共用的資源。您也可以使用 AWS Resource Access Manager (AWS RAM) 主控台來檢視與您的帳戶共用的資源，以及使用具名資源方法與其他 AWS 帳戶共用的資源。

> **Important**
>
> 當有人使用具名資源方法將 Data Catalog 資源的跨帳戶許可授予您的帳戶或 AWS 組織時，Lake Formation 會使用 AWS Resource Access Manager (AWS RAM) 服務來共用資源。如果您的帳戶與授予帳戶位於同一個 AWS 組織中，則共用資源可立即供您使用。
> 不過，如果您的帳戶不在同一個組織中， 會向您的帳戶 AWS RAM 傳送接受或拒絕資源共享的邀請。然後，若要讓共用資源可用，您帳戶中的資料湖管理員必須使用 AWS RAM 主控台或 CLI 接受邀請。
> 如果有 AWS RAM 資源共享邀請等待接受，Lake Formation 主控台會顯示提醒。只有獲授權檢視 AWS RAM 邀請的使用者才會收到提醒。

另請參閱:

- 跨 AWS 帳戶共用 Data Catalog 資料表和資料庫
- Lake Formation 中的跨帳戶資料共用
- 存取共用資料表的基礎資料
- 中繼資料存取控制 （ 如需具名資源方法與用於共用資源的 LF-TBAC 方法的相關資訊。）

主題

- 從 接受資源共享邀請 AWS RAM
- 檢視共用的 Data Catalog 資料表和資料庫

從 接受資源共享邀請 AWS RAM

如果與 AWS 您的帳戶共用 Data Catalog 資源，且您的帳戶與共用帳戶不在同一個 AWS 組織中，則在您接受來自 AWS Resource Access Manager () 的資源共用邀請之前，無法存取共用資源AWS RAM。身為資料湖管理員，您必須先查詢 AWS RAM 待定邀請，然後接受邀請。

您可以使用 AWS RAM 主控台、API 或 AWS Command Line Interface (AWS CLI) 來檢視和接受邀請。

檢視和接受來自 AWS RAM （ 主控台 ） 的資源共享邀請

1. 確保您具有檢視和接受資源共享邀請的必要 AWS Identity and Access Management (IAM) 許可。

   如需資料湖管理員的建議 IAM 政策的相關資訊，請參閱 the section called “Data Lake 管理員許可”。
2. 請遵循AWS RAM 《 使用者指南》中的接受和拒絕邀請中的指示。

從 AWS RAM (AWS CLI) 檢視和接受資源共享邀請

1. 確保您具有檢視和接受資源共享邀請的必要 AWS Identity and Access Management (IAM) 許可。

   如需資料湖管理員的建議 IAM 政策的相關資訊，請參閱 the section called “Data Lake 管理員許可”。
2. 輸入下列命令以檢視擱置中的資源共用邀請。

```
aws ram get-resource-share-invitations
```

輸出格式應類似以下內容。

```
{
    "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"
        }
    ]
}
```

請注意 的狀態PENDING。

3. 將resourceShareInvitationArn金鑰的值複製到剪貼簿。
4. 將值貼到下列命令中，取代 *<invitation-arn>*，然後輸入命令。

```
aws ram accept-resource-share-invitation --resource-share-invitation-
arn <invitation-arn>
```

輸出格式應類似以下內容。

```
{
    "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"
        }
    ]
}
```

請注意 的狀態ACCEPTED。

檢視共用的 Data Catalog 資料表和資料庫

您可以使用 Lake Formation 主控台或 AWS CLI 來檢視與您的帳戶共用的資源。您也可以使用 AWS Resource Access Manager (AWS RAM) 主控台或 CLI 來檢視與您的帳戶共用的資源，以及您與其他 AWS 帳戶共用的資源。

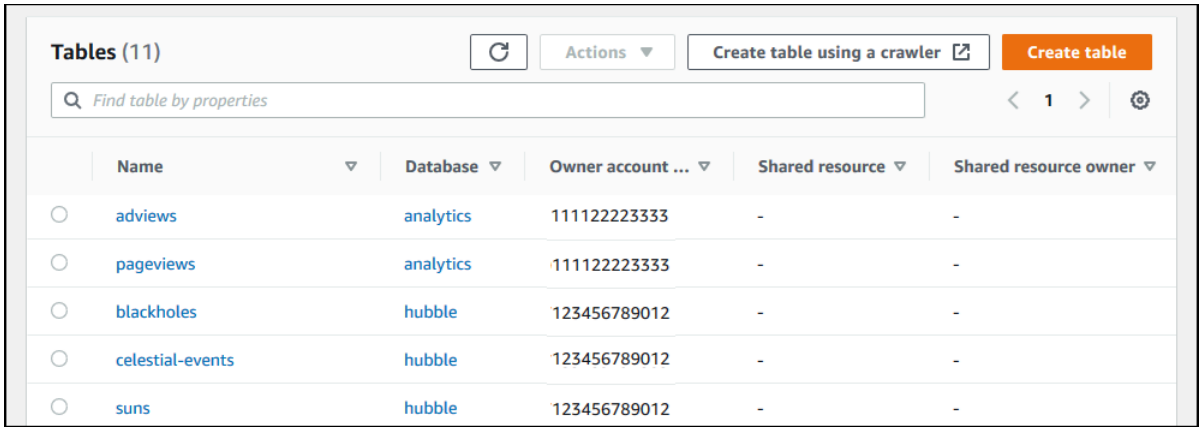
使用 Lake Formation 主控台檢視共用資源

1. 開啟 Lake Formation 主控台，網址為 https://console.aws.amazon.com/lakeformation/。

   以資料湖管理員或已授予共用資料表許可的使用者身分登入。
2. 若要檢視與 AWS 您的帳戶共用的資源，請執行下列其中一項操作：

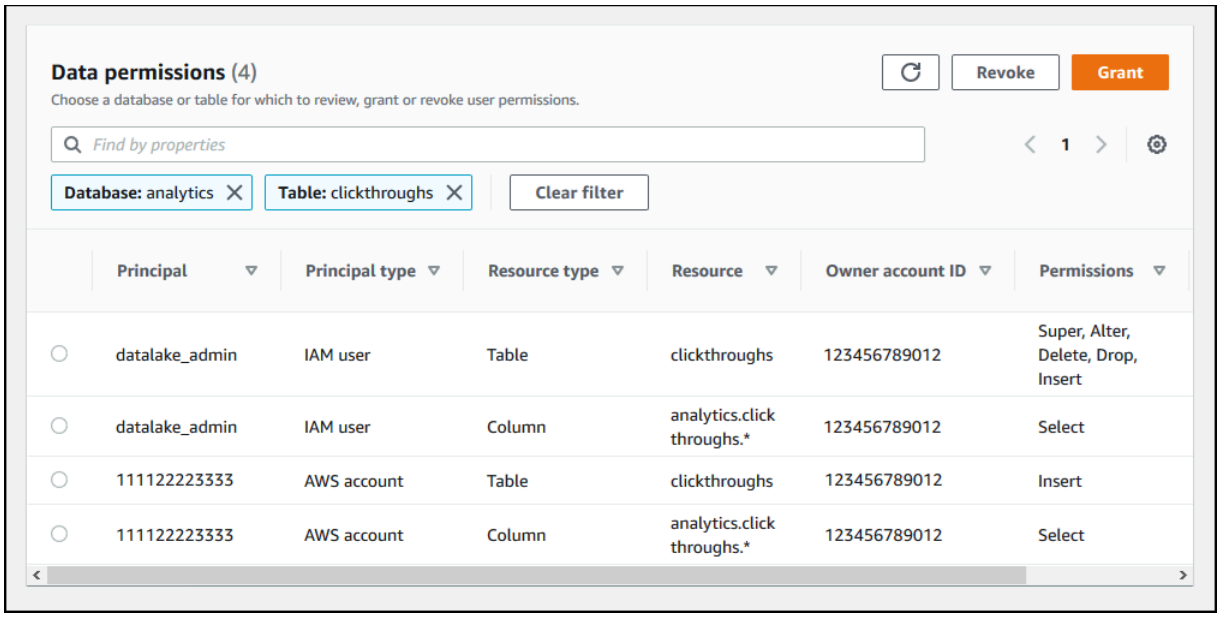
   - 若要檢視與您的帳戶共用的資料表，請在導覽窗格中選擇資料表。
   - 若要檢視與您的帳戶共用的資料庫，請在導覽窗格中選擇資料庫。

   主控台會顯示您帳戶中並與您的帳戶共用的資料庫或資料表清單。對於與您的帳戶共用的資源，主控台會在擁有者 AWS 帳戶 ID 欄 （以下螢幕擷取畫面中的第三個欄） 下方顯示擁有者的帳戶 ID。



3. 若要檢視您與其他 AWS 帳戶或組織共用的資源，請在導覽窗格中選擇資料許可。

   您共用的資源會列在資料許可頁面上，外部帳戶號碼顯示在委託人欄中，如下圖所示。



使用 AWS RAM 主控台檢視共用資源

1. 請確定您具有使用 檢視共用資源所需的 AWS Identity and Access Management (IAM) 許可 AWS RAM。

   您至少必須擁有 `ram:ListResources`許可。此許可包含在 AWS 受管政策中`AWSLakeFormationCrossAccountManager`。

2. 登入 AWS Management Console ，並在 https：//https://console.aws.amazon.com/ram/home 開啟 AWS RAM 主控台。

3. 執行以下任意一項：

   - 若要查看您共用的資源，請在導覽窗格中的我共用下，選擇共用資源。
   - 若要查看與您共用的資源，請在導覽窗格中的與我共用下，選擇共用資源。

建立資源連結

資源連結是 Data Catalog 物件，其為中繼資料資料庫和資料表的連結，通常是來自其他 AWS 帳戶的共用資料庫和資料表。它們有助於跨所有 AWS 區域啟用資料湖中資料的跨帳戶存取。

> **Note**
>
> Lake Formation 支援跨 AWS 區域查詢 Data Catalog 資料表。您可以在指向不同 AWS 區域中共用資料庫和資料表的區域中建立資源連結，以從任何區域存取 Data Catalog 資料庫和資料表。

主題

- 資源連結在 Lake Formation 中如何運作
- 建立共用 Data Catalog 資料表的資源連結
- 建立共用 Data Catalog 資料庫的資源連結
- AWS Glue APIs中的資源連結處理

資源連結在 Lake Formation 中如何運作

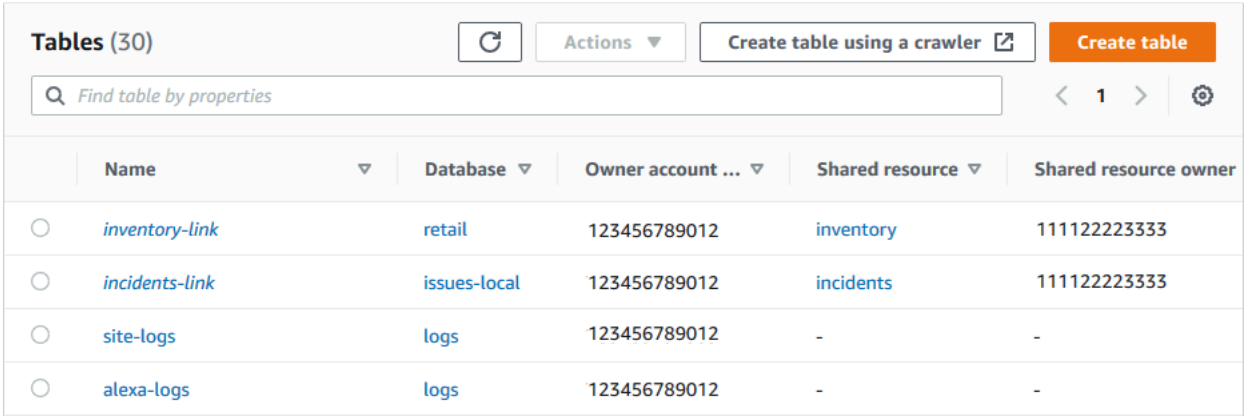
資源連結是 Data Catalog 物件，它是本機或共用資料庫或資料表的連結。建立資料庫或資料表的資源連結後，您可以在任何使用資料庫或資料表名稱的地方使用資源連結名稱。除了您擁有的資料表或與您共用的資料表之外，資料表資源連結也會由 傳回`glue:GetTables()`，並在 Lake Formation 主控台的資料表頁面上顯示為項目。資料庫的資源連結會以類似的方式運作。

建立資料庫或資料表的資源連結可讓您執行下列動作：

- 將不同的名稱指派給 Data Catalog 中的資料庫或資料表。如果不同 AWS 帳戶共用相同名稱的資料庫或資料表，或您帳戶中的多個資料庫具有相同名稱的資料表，這會特別有用。
- 在指向另一個 AWS 區域的資料庫和資料表的區域中建立資源連結，以從任何區域存取 Data Catalog 資料庫和資料表。您可以使用 Athena、Amazon EMR 在任何區域中使用這些資源連結執行查詢，並執行 AWS Glue ETL Spark 任務，而無需複製來源資料或 Glue Data Catalog 中的中繼資料。
- 使用 Amazon Athena 和 Amazon Redshift Spectrum 等整合 AWS 服務來執行存取共用資料庫或資料表的查詢。有些整合服務無法直接存取跨帳戶的資料庫或資料表。不過，他們可以存取您帳戶中的資源連結到其他帳戶中的資料庫和資料表。

> **Note**
>
> 您不需要建立資源連結，即可在AWS Glue擷取、轉換和載入 (ETL) 指令碼中參考共用資料庫或資料表。不過，若要避免多個 AWS 帳戶共用具有相同名稱的資料庫或資料表時發生模棱兩可的情況，您可以在叫用 ETL 操作時建立和使用資源連結或指定目錄 ID。

下列範例顯示 Lake Formation 主控台資料表頁面，其中列出兩個資源連結。資源連結名稱一律以斜體顯示。每個資源連結都會與其連結的共用資源的名稱和擁有者一起顯示。在此範例中， AWS 帳戶 1111-2222-3333 中的資料湖管理員與帳戶 1234-5678-9012 共用 `inventory`和 `incidents`資料表。該帳戶中的使用者接著建立了這些共用資料表的資源連結。



以下是資源連結的備註和限制：

- 需要資源連結才能啟用整合服務，例如 Athena 和 Redshift Spectrum，以查詢共用資料表的基礎資料。這些整合服務中的查詢會根據資源連結名稱建構。
- 假設此設定針對包含資料庫的新資料表只使用 IAM 存取控制已關閉，則只有建立資源連結的主體才能檢視和存取它。若要讓帳戶中的其他主體能夠存取資源連結，請授予其DESCRIBE許可。若要讓其他人捨棄資源連結，請授予其DROP許可。資料湖管理員可以存取帳戶中的所有資源連結。若要捨棄另一個主體建立的資源連結，資料湖管理員必須先授予自己資源連結的DROP許可。如需詳細資訊，請參閱Lake Formation 許可參考。

> **Important**
>
> 授予資源連結的許可不會授予目標 （連結） 資料庫或資料表的許可。您必須分別授予目標的許可。

- 若要建立資源連結，您需要 Lake Formation CREATE_TABLE或 CREATE_DATABASE 許可，以及 glue:CreateTable或 glue:CreateDatabase AWS Identity and Access Management (IAM) 許可。
- 您可以建立本機 （ 擁有 ） Data Catalog 資源的資源連結，以及與 AWS 您的帳戶共用的資源連結。
- 當您建立資源連結時，不會執行檢查以查看目標共用資源是否存在，或您對資源是否具有跨帳戶許可。這可讓您以任何順序建立資源連結和共用資源。
- 如果您刪除資源連結，則不會捨棄連結的共用資源。如果您捨棄共用資源，則不會刪除該資源的資源連結。
- 您可以建立資源連結鏈。不過，這樣做沒有價值，因為 APIs只遵循第一個資源連結。

另請參閱：

- 授予 Data Catalog 資源的許可

建立共用 Data Catalog 資料表的資源連結

您可以使用 AWS Lake Formation 主控台、API 或 AWS Command Line Interface ()，在任何 AWS 區域中建立共用資料表的資源連結AWS CLI。

建立共用資料表的資源連結 （ 主控台 ）

1. 在 https：//https://console.aws.amazon.com/lakeformation/ 開啟 AWS Lake Formation 主控台。以擁有資料庫 Lake Formation CREATE_TABLE許可的主體身分登入，以包含資源連結。
2. 在導覽窗格中，選擇 Data Catalog 下的資料表，然後選擇建立、資源連結。
3. 在建立資源連結頁面上，提供下列資訊：

   資源連結名稱

   輸入遵循與資料表名稱相同規則的名稱。名稱可以與目標共用資料表相同。

   資料庫

   本機 Data Catalog 中要包含資源連結的資料庫。

   共用資料表擁有者區域

   如果您要在不同區域中建立資源連結，請選取目標共用資料表的區域。

共用資料表

從清單中選擇共用資料表，或輸入本機（擁有）或共用資料表名稱。

此清單包含與您的帳戶共用的所有資料表。請注意與每個資料表一起列出的資料庫和擁有者帳戶 ID。如果您沒有看到您知道已與 帳戶共用的資料表，請檢查下列項目：

- 如果您不是資料湖管理員，請檢查資料湖管理員是否授予資料表上的 Lake Formation 許可。
- 如果您是資料湖管理員，而且您的帳戶與授予帳戶不在同一個 AWS 組織中，請確定您已接受資料表的 AWS Resource Access Manager (AWS RAM) 資源共享邀請。如需詳細資訊，請參閱從 接受資源共享邀請 AWS RAM。

共用資料表的資料庫

如果您從清單中選取共用資料表，則此欄位會填入外部帳戶中共用資料表的資料庫。否則，請在外部帳戶中輸入本機資料庫（用於本機資料表的資源連結）或共用資料表的資料庫。

共用資料表擁有者

如果您從清單中選取共用資料表，則此欄位會填入共用資料表的擁有者帳戶 ID。否則，請輸入 AWS 您的帳戶 ID（適用於本機資料表的資源連結）或共用資料表 AWS 的帳戶 ID。

4. 選擇建立以建立資源連結。

然後，您可以在資料表頁面的名稱欄下檢視資源連結名稱。

5. （選用）將資源連結的 Lake Formation DESCRIBE許可授予必須能夠檢視連結和存取目標資料表的主體。

不過，授予資源連結的許可並不會授予目標（連結）資料庫或資料表的許可。您必須分別授予目標資料庫的許可，才能在 Athena 中顯示資料表/資源連結。

若要在相同區域中建立共用資料表的資源連結 (AWS CLI)

1. 輸入與以下相似的命令。

```
aws glue create-table --database-name myissues --table-input
 '{"Name":"my_customers","TargetTable":
{"CatalogId":"111122223333","DatabaseName":"issues","Name":"customers"}}'
```

此命令會建立名為my_customers共用資料表 的資源連結customers，該資料表位於帳戶 1111-2222-3333 issues的 AWS 資料庫中。資源連結會存放在本機資料庫 中myissues。

2. （選用）將資源連結的 Lake Formation DESCRIBE許可授予必須能夠檢視連結和存取目標資料表的主體。

   不過，授予資源連結的許可並不會授予目標（連結）資料表的許可。您必須分別授予目標資料庫的許可，才能在 Athena 中顯示資料表/資源連結。

若要在不同區域中建立共用資料表的資源連結 (AWS CLI)

1. 輸入與以下相似的命令。

```
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"
        }
    }
}‘
```

   此命令`rl_useast1salestb_ireland`會在歐洲（愛爾蘭）區域中建立名為 的資源連結至共用資料表 `useast1_salestb`，該資料表在美國東部（維吉尼亞北部）區域中`useast1_salesdb` AWS 帳戶 444455556666 的資料庫中。資源連結會存放在本機資料庫 中`ireland_db`。

2. 將 Lake Formation DESCRIBE許可授予必須能夠檢視連結並透過連結存取連結目標的主體。

   不過，授予資源連結的許可並不會授予目標（連結）資料表的許可。您必須分別授予目標資料表的許可，才能在 Athena 中顯示資料表/資源連結。

> 另請參閱：
>
> - 資源連結在 Lake Formation 中如何運作
> - DESCRIBE

建立共用 Data Catalog 資料庫的資源連結

您可以使用 AWS Lake Formation 主控台、API 或 AWS Command Line Interface () 建立共用資料庫的資源連結AWS CLI。

建立共用資料庫的資源連結 （主控台）

1. 在 https：//https://console.aws.amazon.com/lakeformation/ 開啟 AWS Lake Formation 主控台。以資料湖管理員或資料庫建立者身分登入。

   資料庫建立者是已獲授予 Lake Formation CREATE_DATABASE許可的主體。
2. 在導覽窗格中，選擇資料庫，然後選擇建立、資源連結。
3. 在建立資源連結頁面上，提供下列資訊：

   資源連結名稱

   輸入與資料庫名稱遵守相同規則的名稱。名稱可以與目標共用資料庫相同。

   目的地目錄

   選取資料庫資源連結的目的地目錄。

   共用資料庫擁有者區域

   如果您要在不同區域中建立資源連結，請選取目標共用資料庫的區域。

   共用資料庫

   從清單中選擇資料庫，或輸入本機 （擁有） 或共用資料庫名稱。

   此清單包含與您的帳戶共用的所有資料庫。請注意每個資料庫列出的擁有者帳戶 ID。如果您看不到已知已與 帳戶共用的資料庫，請檢查下列項目：

   - 如果您不是資料湖管理員，請檢查資料湖管理員是否已授予資料庫上的 Lake Formation 許可。
   - 如果您是資料湖管理員，而且您的帳戶與授予帳戶不在同一個 AWS 組織中，請確定您已接受資料庫的 AWS Resource Access Manager (AWS RAM) 資源共享邀請。如需詳細資訊，請參閱從 接受資源共享邀請 AWS RAM。

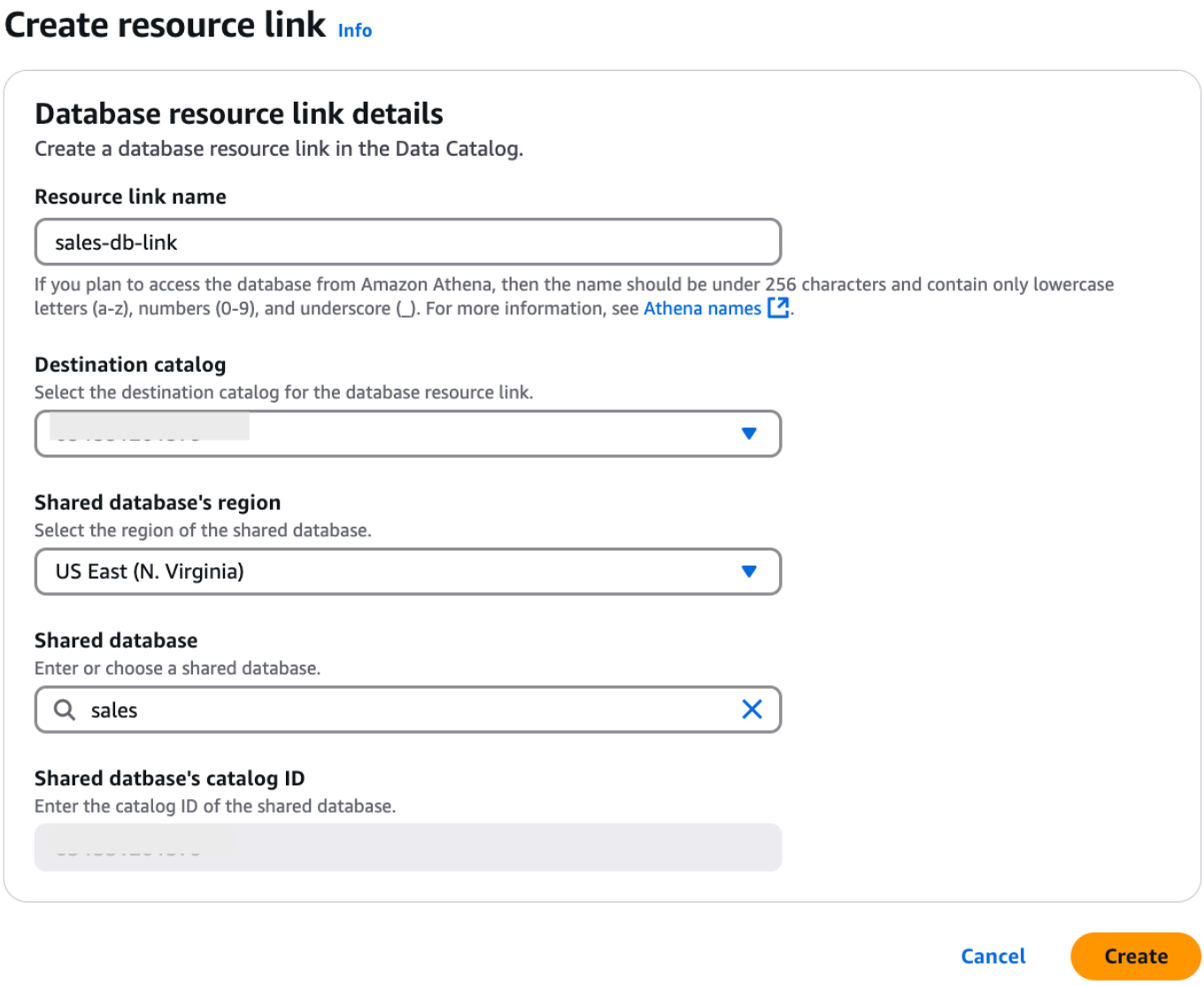
   共用資料庫擁有者

   如果您從清單中選取共用資料庫，則此欄位會填入共用資料庫的擁有者帳戶 ID。否則，請輸入 AWS 您的帳戶 ID （適用於本機資料庫的資源連結） 或共用資料庫 AWS 的帳戶 ID。

共用資料庫的目錄 ID

輸入共用資料庫的目錄 ID。建立從另一個 AWS 帳戶共用之 databse 的資源連結時，您需要指定此目錄 ID，以識別哪個帳戶的 Data Catalog 包含來源資料庫。

當您從下拉式功能表中選取共用資料庫時，系統會自動填入擁有該資料庫且已與您共用該資料庫之帳戶的目錄 ID。



4. 選擇建立以建立資源連結。

   然後，您可以在資料庫頁面的名稱欄下檢視資源連結名稱。

5. （選用）將資源連結的 Lake Formation DESCRIBE許可授予必須能夠檢視連結和存取目標資料庫的歐洲（愛爾蘭）區域的主體。

不過，授予資源連結的許可並不會授予目標（連結）資料庫或資料表的許可。您必須分別授予目標資料庫的許可，才能在 Athena 中顯示資料表/資源連結。

若要在相同區域中建立共用資料庫的資源連結 (AWS CLI)

1. 輸入與以下相似的命令。

```
aws glue create-database --database-input '{"Name":"myissues","TargetDatabase":
{"CatalogId":"111122223333","DatabaseName":"issues"}}'
```

此命令會建立名為 的資源連結myissues至共用資料庫 issues，該資料庫位於 AWS 帳戶 1111-2222-3333 中。

2. （選用）將 Lake Formation DESCRIBE許可授予資源連結上必須能夠檢視連結和存取目標資料庫或資料表的主體。

不過，授予資源連結的許可並不會授予目標（連結）資料庫或資料表的許可。您必須分別授予目標資料庫的許可，才能在 Athena 中顯示資料表/資源連結。

若要在不同區域中建立共用資料庫的資源連結 (AWS CLI)

1. 輸入與以下相似的命令。

```
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"
       }
    }
}'
```

此命令會建立在歐洲（愛爾蘭）區域中 AWS 帳戶 111122223333 rl_useast1shared_irelanddb中名為 的資源連結至共用資料庫 useast1shared_db，該資料庫位於美國東部（維吉尼亞北部）區域中 AWS 的帳戶 444455556666。

2. 將 Lake Formation DESCRIBE許可授予來自歐洲 （愛爾蘭） 區域的主體，這些主體必須能夠檢視連結並透過連結存取連結目標。

> 另請參閱：
>
> - 資源連結在 Lake Formation 中如何運作
> - DESCRIBE

AWS Glue APIs中的資源連結處理

下表說明 AWS Glue Data Catalog APIs如何處理資料庫和資料表資源連結。對於所有 Get* API 操作，只有發起人在 上具有許可的資料庫和資料表才會傳回。此外，透過資源連結存取目標資料庫或資料表時，您必須同時擁有目標和資源連結的 AWS Identity and Access Management (IAM) 和 Lake Formation 許可。資源連結上所需的 Lake Formation 許可為 DESCRIBE。如需詳細資訊，請參閱DESCRIBE。

資料庫 API 操作

API 操作	資源連結處理
CreateDatabase	如果資料庫是資源連結， 會建立指定目標資料庫的資源連結。
UpdateDatabase	如果指定的資料庫是資源連結， 會遵循連結並更新目標資料庫。如果必須修改資源連結以連結至不同的資料庫，您必須將其刪除並建立新的資料庫。
DeleteDatabase	刪除資源連結。它不會刪除連結的 （目標） 資料庫。
GetDatabase	如果發起人對目標具有許可， 會遵循連結來傳回目標的屬性。否則，它會傳回連結的屬性。
GetDatabases	傳回資料庫清單，包括資源連結。對於結果集中的每個資源連結，操作會遵循連結來取得連結目標的屬性。您必須指定 ResourceShareType = ALL才能查看與您的帳戶共用的資料庫。

資料表 API 操作

API 操作	資源連結處理
CreateTable	如果資料庫是資源連結， 會遵循資料庫連結，並在目標資料庫中建立資料表。如果資料表是資源連結，操作會在指定的資料庫中建立資源連結。不支援透過資料庫資源連結建立資料表資源連結。
UpdateTable	如果資料表或指定的資料庫是資源連結， 會更新目標資料表。如果資料表和資料庫都是資源連結，則操作會失敗。
DeleteTable	如果指定的資料庫是資源連結， 會遵循連結，並刪除目標資料庫中的資料表或資料表資源連結。如果資料表是資源連結，操作會刪除指定資料庫中的資料表資源連結。刪除資料表資源連結不會刪除目標資料表。
BatchDeleteTable	與 DeleteTable 相同。
GetTable	如果指定的資料庫是資源連結， 會遵循資料庫連結，並從目標資料庫傳回資料表或資料表資源連結。否則，如果資料表是資源連結，操作會遵循連結並傳回目標資料表屬性。
GetTables	如果指定的資料庫是資源連結， 會遵循資料庫連結，並從目標資料庫傳回資料表和資料表資源連結。如果目標資料庫是來自另一個 AWS 帳戶的共用資料庫，則操作只會傳回該資料庫中的共用資料表。它不會遵循目標資料庫中的資料表資源連結。否則，如果指定的資料庫是本機 （擁有） 資料庫，操作會傳回本機資料庫中的所有資料表，並遵循每個資料表資源連結來傳回目標資料表屬性。
SearchTables	傳回資料表和資料表資源連結。它不會遵循連結來傳回目標資料表屬性。您必須指定 ResourceShareType = ALL才能查看與您的帳戶共用的資料表。
GetTableVersion	與 GetTable 相同。
GetTableVersions	與 GetTable 相同。
DeleteTableVersion	與 DeleteTable 相同。

API 操作	資源連結處理
BatchDeleteTableVersion	與 DeleteTable 相同。

分割區 API 操作

API 操作	資源連結處理
CreatePartition	如果指定的資料庫是資源連結，會遵循資料庫連結，並在目標資料庫中的指定資料表中建立分割區。如果資料表是資源連結，則操作會遵循資源連結，並在目標資料表中建立分割區。不支援透過資料表資源連結和資料庫資源連結建立分割區。
BatchCreatePartition	與 CreatePartition 相同。
UpdatePartition	如果指定的資料庫是資源連結，會遵循資料庫連結，並在目標資料庫中的指定資料表中更新分割區。如果資料表是資源連結，則操作會遵循資源連結，並更新目標資料表中的分割區。不支援透過資料表資源連結和資料庫資源連結更新分割區。
DeletePartition	如果指定的資料庫是資源連結，會遵循資料庫連結，並在目標資料庫中的指定資料表中刪除分割區。如果資料表是資源連結，操作會遵循資源連結，並刪除目標資料表中的分割區。不支援透過資料表資源連結和資料庫資源連結刪除分割區。
BatchDeletePartition	與 DeletePartition 相同。
GetPartition	如果指定的資料庫是資源連結，會遵循資料庫連結，並從指定的資料表傳回分割區資訊。否則，如果資料表是資源連結，則操作會遵循連結並傳回分割區資訊。如果資料表和資料庫都是資源連結，則會傳回空的結果集。
GetPartitions	如果指定的資料庫是資源連結，會遵循資料庫連結，並傳回指定資料表中所有分割區的分割區資訊。否則，如果資料表是資源連結，則

API 操作	資源連結處理
	操作會遵循連結並傳回分割區資訊。如果資料表和資料庫都是資源連結，則會傳回空的結果集。
`BatchGetPartition`	與 `GetPartition` 相同。

使用者定義的函數 API 操作

API 操作	資源連結處理
（所有 API 操作）	如果資料庫是資源連結， 會遵循資源連結，並在目標資料庫上執行操作。

另請參閱：

- 資源連結在 Lake Formation 中如何運作

跨區域存取資料表

Lake Formation 支援跨 AWS 區域查詢 Data Catalog 資料表。您可以使用 Amazon Athena、Amazon EMR 和 AWS Glue ETL，在指向來源資料庫和資料表的其他區域中建立資源連結，從其他區域存取區域中的資料。使用跨區域資料表存取，您可以跨區域存取資料，而無需將基礎資料或中繼資料複製到 Data Catalog。

例如， 您可以在生產者帳戶中與區域 A 中的取用者帳戶共用資料庫或資料表。接受區域 A 中的資源共用邀請後， 消費者帳戶的資料湖管理員可以建立區域 B 中共用資源的資源連結。消費者帳戶管理員可以將共用資源的許可授予區域 A 中該帳戶中的 IAM 主體，並授予區域 B 中的資源連結許可。使用資源連結， 消費者帳戶中的主體可以查詢來自區域 B 的共用資料。

您也可以在生產者帳戶中的區域 A 中託管 Amazon S3 資料來源，並在區域 B 的中央帳戶中註冊資料位置。您可以在中央帳戶中建立 Data Catalog 資源、設定 Lake Formation 許可，以及與帳戶中的取用者或區域 B 的外部帳戶共用資料。跨區域功能可讓使用者使用資源連結從區域 C 存取這些資料目錄資料表。

使用此功能，您可以在跨 區域的 Apache Hive 中繼存放區中查詢聯合資料庫，也可以在執行查詢時將本機區域中的資料表與另一個區域中的資料表聯結。

Lake Formation 支援具有跨區域資料表存取的下列功能：

- LF 標籤型存取控制
- 精細存取控制許可
- 在具有適當許可的共用資料庫或資料表上寫入操作
- 在帳戶層級跨帳戶資料共用，並直接與 IAM 主體層級共用

具有 `Create_Database`和 `Create_Table`許可的非管理使用者可以建立跨區域資源連結。

> **Note**
>
> 您可以在任何區域中建立跨區域資源連結並存取資料，而無需套用 Lake Formation 許可。對於未向 Lake Formation 註冊的 Amazon S3 中的來源資料，存取權取決於 Amazon S3 和 AWS Glue 動作的 IAM 許可政策。

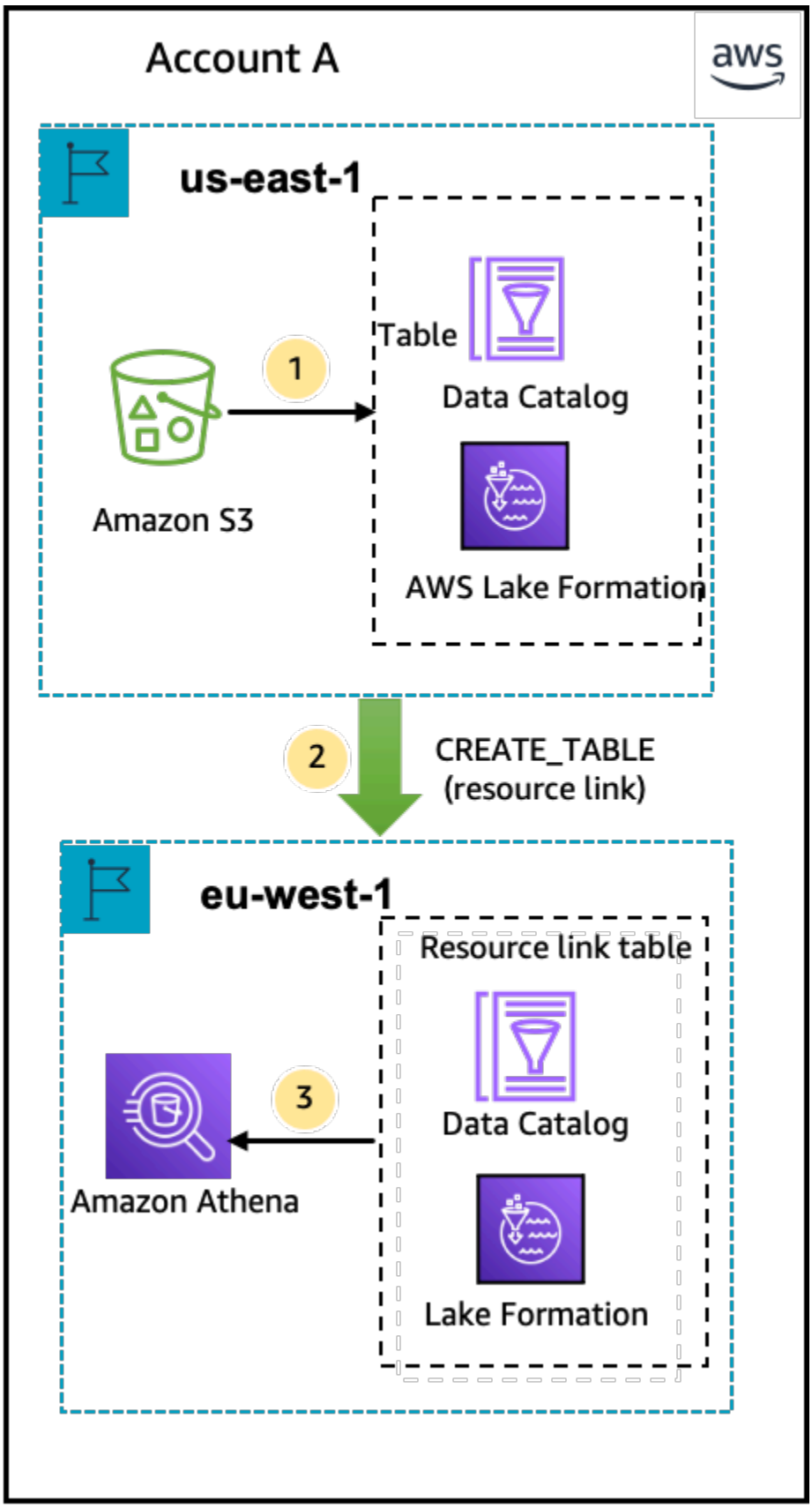
如需限制的詳細資訊，請參閱跨區域資料存取限制。

工作流程

下圖顯示從相同 AWS 帳戶和外部帳戶跨 AWS 區域存取資料的工作流程。

存取相同 AWS 帳戶中共用資料表的工作流程

在下圖中，資料會與美國東部（維吉尼亞北部）區域中相同 AWS 帳戶中的使用者共用，而使用者會從歐洲（愛爾蘭）區域查詢共用資料。


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

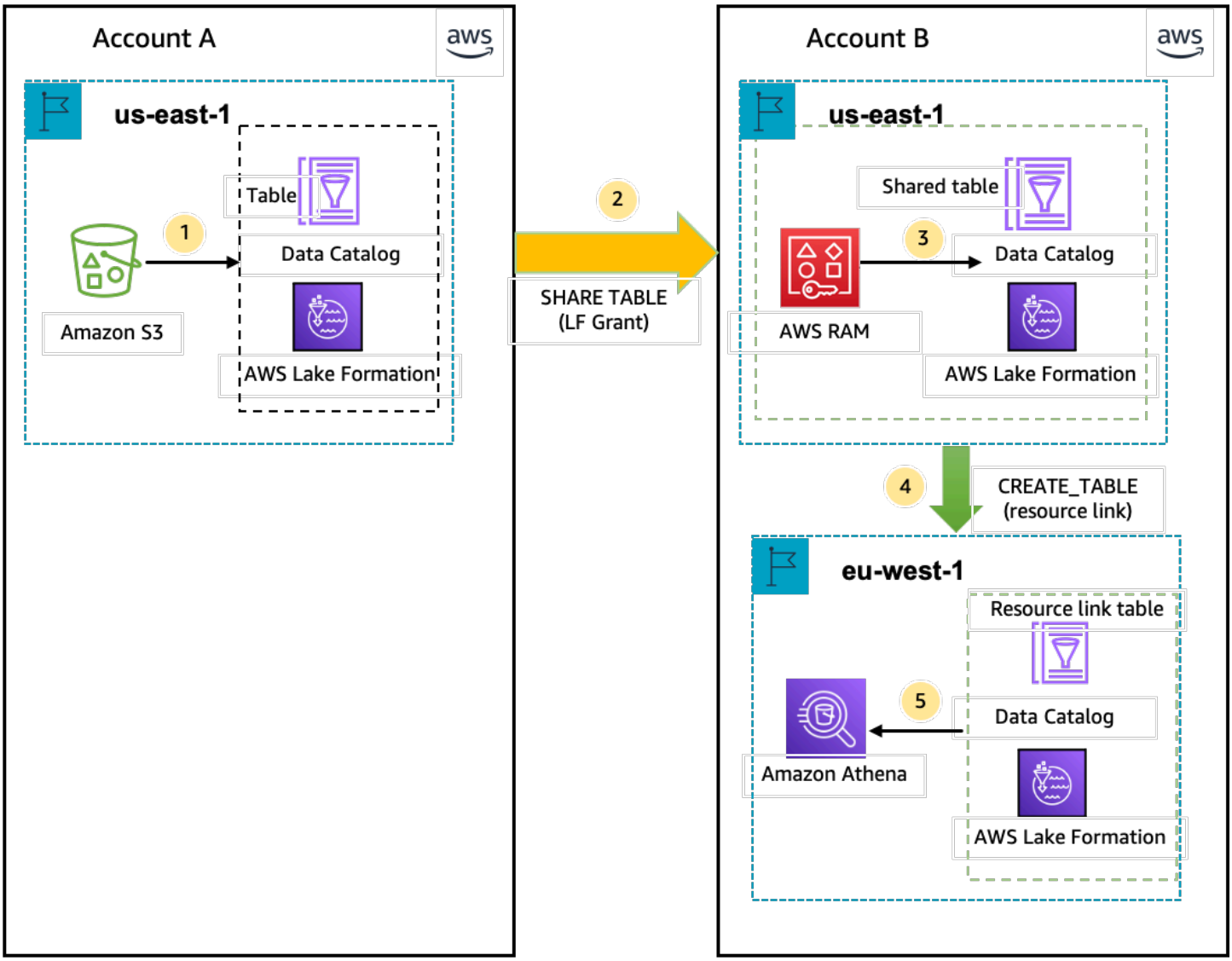
資料湖管理員會執行下列活動（步驟 1-2)：

1. 資料湖管理員使用 Data Catalog 資料庫和資料表設定 AWS 帳戶，並向美國東部（維吉尼亞北部）區域的 Lake Formation 註冊 Amazon S3 資料位置。

   將 Data Catalog 資源（圖表中的產品資料表）的`Select`許可授予相同帳戶中的委託人（使用者）。
2. 在歐洲（愛爾蘭）區域中建立指向美國東部（維吉尼亞北部）區域中來源資料表的資源連結。`DESCRIBE` 准許從歐洲（愛爾蘭）區域到委託人的資源連結。
3. 使用者使用 Athena 從歐洲（愛爾蘭）區域查詢資料表。

存取與外部 AWS 帳戶共用之資料表的工作流程

在下圖中，生產者帳戶（帳戶 A) 託管 Amazon S3 儲存貯體、註冊資料位置，以及與美國東部（維吉尼亞北部）區域中的消費者帳戶（帳戶 B) 和來自消費者帳戶（帳戶 B) 的使用者共用資料目錄資料表，查詢來自歐洲（愛爾蘭）區域的資料表。



1. 資料湖管理員使用 Data Catalog 資源設定 AWS 帳戶（生產者帳戶），以及在美國東部（維吉尼亞北部）區域向 Lake Formation 註冊的 Amazon S3 資料位置。
2. 生產者帳戶的資料湖管理員將 Data Catalog 資料表分享給取用者帳戶。
3. 消費者帳戶的資料湖管理員接受美國東部（維吉尼亞北部）區域中的資料共用邀請，並將共用資料表的`Select`許可授予相同區域的委託人。
4. 消費者帳戶的資料湖管理員會在歐洲（愛爾蘭）區域中建立指向美國東部（維吉尼亞北部）區域中目標共用資料表的資源連結，並授予來自歐洲（愛爾蘭）區域之資源連結的使用者`DESCRIBE`許可。
5. 使用者使用 Athena 查詢來自歐洲（愛爾蘭）區域的資料。

設定跨區域資料表存取

若要從不同區域存取資料，您必須先在註冊 Amazon S3 資料位置的區域中設定 Data Catalog 資料庫和資料表。您可以與帳戶中的主體或其他帳戶中的主體共用 Data Catalog 資料庫和資料表。然後，您需要建立資料湖管理員，他們可以建立指向使用者查詢資料之區域中目標共用資料位置的資源連結。

查詢來自不同 區域的相同 帳戶中共用的資料

在本節中，目標共用資料表區域稱為區域 A，使用者從區域 B 執行查詢。

1. 區域 A 中的帳戶設定 （您在其中建立和共用資料）

   資料湖管理員需要完成下列動作：

   a. 註冊 Amazon S3 資料位置。

      如需詳細資訊，請參閱將 Amazon S3 位置新增至您的資料湖。

   b. 在帳戶中建立資料庫和資料表。這也可以由具有建立資料庫和資料表許可的非管理使用者完成。

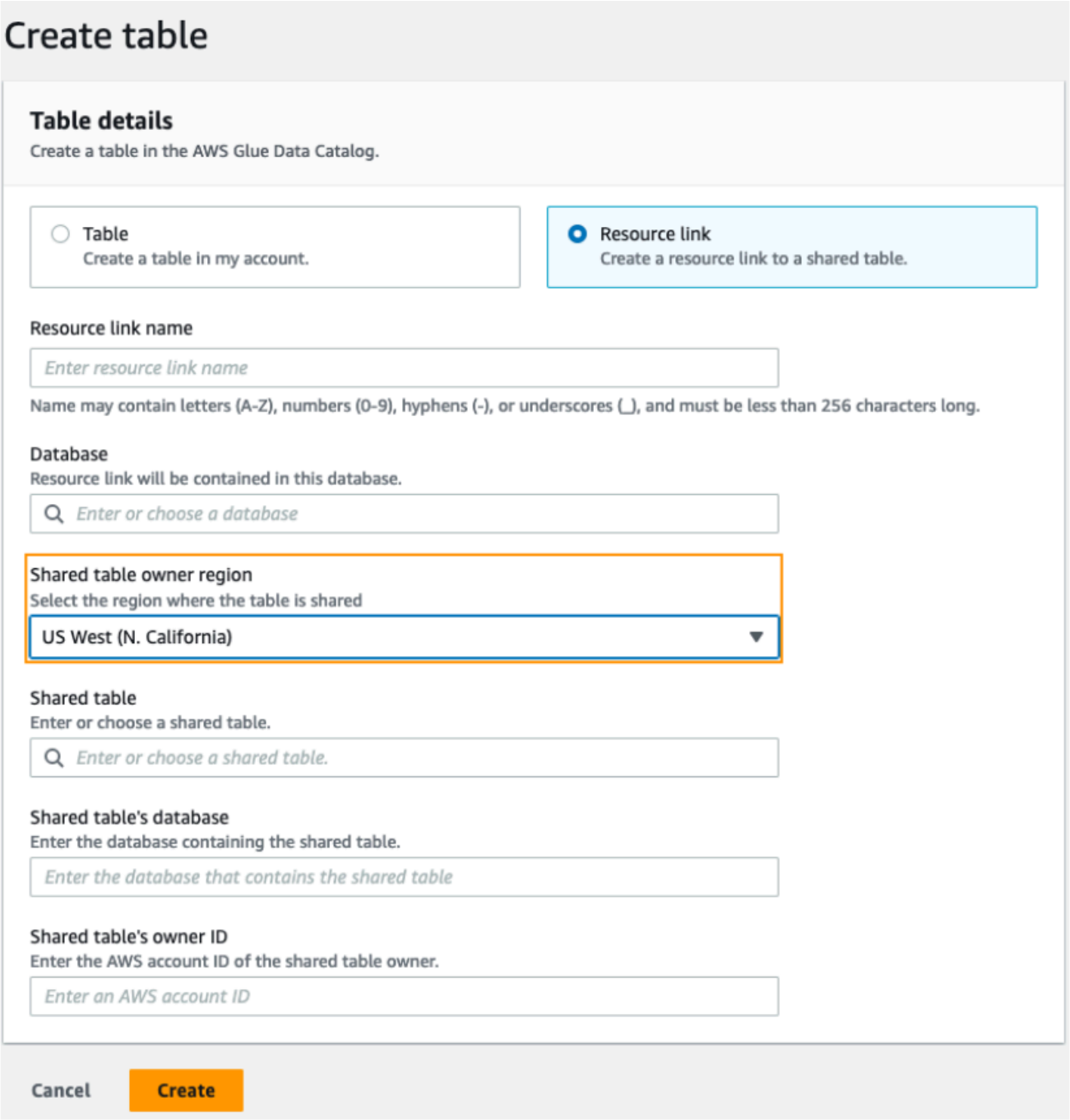
   c. 使用 將資料表上的資料許可授予委託人`Grantable permissions`。

      如需詳細資訊，請參閱 授予 Data Catalog 資源的許可。

2. 區域 B （您存取資料的位置） 中的帳戶設定

   資料湖管理員需要完成下列動作：

   a. 在區域 B 中建立指向區域 A 中目標共用資料表的資源連結。在建立資料表畫面上指定共用資料表擁有者區域。



如需建立資料庫和資料表資源連結的說明，請參閱 建立資源連結。

b. 在區域 B 的資源連結上授予 IAM 主體`Describe`許可。

如需授予資源連結許可的詳細資訊，請參閱 授予資源連結許可。

區域 B 中的 IAM 主體可以使用 Athena 透過連結查詢目標資料表。

從不同區域存取跨帳戶資料

1. 生產者/授權方帳戶設定

   資料湖管理員需要完成下列動作：

   a. 在區域 A 中設定生產者/授權者帳戶。
   b. 在區域 A 中註冊 Amazon S3 資料位置。
   c. 建立資料庫和資料表。這可由具有建立資料表許可的非管理使用者完成。
   d. 使用 將資料許可授予區域 A 中資料表上的消費者/承授者帳戶`Grantable permissions`。

      如需詳細資訊，請參閱從外部帳戶跨 AWS 帳戶 或 IAM 主體共用 Data Catalog 資料表和資料庫。

2. 消費者/承授者帳戶設定

   資料湖管理員需要完成下列動作：

   a. 在區域 A AWS RAM 接受來自 的資源共用邀請。
   b. 在指向共用資料表的區域 B 中建立資源連結。區域 B 是使用者想要查詢資料表的地方。
   c. 將共用資料表上的資料許可授予區域 A 中的 IAM 主體。

      > **Note**
      > 您必須將許可授予共用資料表所在相同區域中的共用資料表。

   d. 授予許可給區域 B 中資源連結的主體。

      然後，區域 B 中消費者帳戶中的主體會使用 Athena 查詢區域 B 的共用資料表。

中的安全性 AWS Lake Formation

的雲端安全 AWS 是最高優先順序。身為 AWS 客戶，您可以受益於資料中心和網路架構，這些架構是為了滿足最安全敏感組織的需求而建置。

安全是 AWS 與您之間的共同責任。共同責任模型 將此描述為雲端的安全和雲端內的安全：

- 雲端的安全性 – AWS 負責保護在 AWS Cloud 中執行 AWS 服務的基礎設施。 AWS 也為您提供可安全使用的服務。在 AWS 合規計畫中，第三方稽核員會定期測試並驗證我們的安全功效。若要了解適用的合規計劃 AWS Lake Formation，請參閱AWS 合規計劃範圍內的服務。
- 雲端的安全性 – 您的責任取決於您使用 AWS 的服務。您也必須對其他因素負責，包括資料的機密性、您公司的要求和適用法律和法規。

本文件可協助您了解如何在使用 Lake Formation 時套用共同責任模型。下列主題說明如何設定 Lake Formation 以符合您的安全與合規目標。您也會了解如何使用其他 AWS 服務來協助您監控和保護 Lake Formation 資源。

主題

- Lake Formation 中的資料保護
- 中的基礎設施安全 AWS Lake Formation
- 預防跨服務混淆代理人
- 安全事件登入 AWS Lake Formation

Lake Formation 中的資料保護

AWS 共同責任模型適用於 AWS Lake Formation 中的資料保護。如此模型所述， AWS 負責保護執行所有 的全域基礎設施 AWS 雲端。您負責維護在此基礎設施上託管內容的控制權。您也同時負責所使用 AWS 服務 的安全組態和管理任務。如需資料隱私權的詳細資訊，請參閱資料隱私權常見問答集。如需有關歐洲資料保護的相關資訊，請參閱 AWS 安全性部落格上的 AWS 共同的責任模型和 GDPR 部落格文章。

基於資料保護目的，建議您保護 AWS 帳戶 登入資料，並使用 AWS IAM Identity Center 或 AWS Identity and Access Management (IAM) 設定個別使用者。如此一來，每個使用者都只會獲得授與完成其任務所必須的許可。我們也建議您採用下列方式保護資料：

- 每個帳戶均要使用多重要素驗證 (MFA)。

- 使用 SSL/TLS 與 AWS 資源通訊。我們需要 TLS 1.2 並建議使用 TLS 1.3。
- 使用 設定 API 和使用者活動記錄 AWS CloudTrail。如需有關使用 CloudTrail 追蹤擷取 AWS 活動的資訊，請參閱AWS CloudTrail 《 使用者指南》中的使用 CloudTrail 追蹤。
- 使用 AWS 加密解決方案，以及其中的所有預設安全控制 AWS 服務。
- 使用進階的受管安全服務 (例如 Amazon Macie)，協助探索和保護儲存在 Amazon S3 的敏感資料。
- 如果您在 AWS 透過命令列界面或 API 存取 時需要 FIPS 140-3 驗證的密碼編譯模組，請使用 FIPS 端點。如需有關 FIPS 和 FIPS 端點的更多相關資訊，請參閱聯邦資訊處理標準 (FIPS) 140-3。

我們強烈建議您絕對不要將客戶的電子郵件地址等機密或敏感資訊，放在標籤或自由格式的文字欄位中，例如名稱欄位。這包括當您使用 Lake Formation 或使用 AWS 服務 主控台、API AWS CLI或其他 AWS SDKs 時。您在標籤或自由格式文字欄位中輸入的任何資料都可能用於計費或診斷日誌。如果您提供外部伺服器的 URL，我們強烈建議請勿在驗證您對該伺服器請求的 URL 中包含憑證資訊。

靜態加密

AWS Lake Formation 支援下列領域的資料加密：

- Amazon Simple Storage Service (Amazon S3) 資料湖中的資料。

  Lake Formation 支援使用 AWS Key Management Service() 進行資料加密AWS KMS。資料通常透過AWS Glue擷取、轉換和載入 (ETL) 任務寫入資料湖。如需有關如何加密AWS Glue任務所寫入資料的資訊，請參閱《 AWS Glue 開發人員指南》中的加密爬蟲程式、任務和開發端點所寫入的資料。
- AWS Glue Data Catalog，這是 Lake Formation 存放描述資料湖中資料的中繼資料表的位置。

  如需詳細資訊，請參閱《 AWS Glue 開發人員指南》中的加密您的資料目錄。

若要將 Amazon S3 位置新增為資料湖中的儲存體，請使用 註冊該位置 AWS Lake Formation。然後，您可以使用 Lake Formation 許可，對指向此位置的 AWS Glue Data Catalog 物件，以及該位置的基礎資料進行精細存取控制。

Lake Formation 支援註冊包含加密資料的 Amazon S3 位置。如需詳細資訊，請參閱註冊加密的 Amazon S3 位置。

中的基礎設施安全 AWS Lake Formation

作為受管服務， AWS Lake Formation 受到 Amazon Web Services：安全程序概觀白皮書中所述的 AWS 全球網路安全程序的保護。

您可以使用 AWS 已發佈的 API 呼叫，透過網路存取 Lake Formation。用戶端必須支援 Transport Layer Security (TLS) 1.0 或更新版本。建議使用 TLS 1.2 或更新版本。用戶端也必須支援具備完美轉送私密 (PFS) 的密碼套件，例如臨時 Diffie-Hellman (DHE) 或橢圓曲線臨時 Diffie-Hellman (ECDHE)。現代系統 (如 Java 7 和更新版本) 大多會支援這些模式。

此外，請求必須使用存取金鑰 ID 和與 IAM 主體相關聯的私密存取金鑰來簽署。或者，您可以使用 AWS Security Token Service (AWS STS) 以產生暫時安全憑證以簽署請求。

預防跨服務混淆代理人

混淆代理人問題屬於安全性問題，其中沒有執行動作許可的實體可以強制具有更多許可的實體執行該動作。在 中 AWS，跨服務模擬可能會導致混淆代理人問題。在某個服務 (呼叫服務) 呼叫另一個服務 (被呼叫服務) 時，可能會發生跨服務模擬。可以操縱呼叫服務來使用其許可，以其不應有存取許可的方式對其他客戶的資源採取動作。為了避免這種情況， AWS 提供工具，協助您保護所有 服務的資料，而服務主體已獲得您帳戶中資源的存取權。

若要限制 AWS Lake Formation 為資源提供另一項服務的許可，我們建議在資源政策中使用 `aws:SourceArn` 和 `aws:SourceAccount` 全域條件內容索引鍵。如果同時使用全域條件內容索引鍵，則在相同政策陳述式中使用 `aws:SourceAccount` 值和 `aws:SourceArn` 值中的帳戶時，必須使用相同的帳戶 ID。

目前，Lake Formation 僅支援`aws:SourceArn`下列格式：

```
arn:aws:lakeformation:aws-region:account-id:*
```

下列範例示範如何使用 Lake Formation 中的 `aws:SourceArn`和 `aws:SourceAccount`全域條件內容索引鍵，以防止混淆代理人問題。

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

安全事件登入 AWS Lake Formation

AWS Lake Formation 已與 服務整合 AWS CloudTrail，此服務提供由使用者、角色或 Lake Formation 中的 AWS 服務所採取之動作的記錄。CloudTrail 會將 Lake Formation 的所有 API 呼叫擷取為事件。擷取的呼叫包括來自 Lake Formation 主控台的呼叫 AWS Command Line Interface、 和對 Lake Formation API 操作的程式碼呼叫。

如需 Lake Formation 中事件記錄的詳細資訊，請參閱 使用 記錄 AWS Lake Formation API 呼叫 AWS CloudTrail。

> **Note**
>
> `GetTableObjects`、 `UpdateTableObjects`和 `GetWorkUnitResults`是大量資料平面操作。這些 APIs呼叫目前不會記錄到 CloudTrail。如需 CloudTrail 中資料平面操作的詳細資訊，請參閱AWS CloudTrail 《 使用者指南》中的記錄線索的資料事件。
> Lake Formation 中支援其他 CloudTrail 事件的變更將記載於 的文件歷史記錄 AWS Lake Formation。

將第三方服務與 整合 Lake Formation

與 整合AWS Lake Formation可讓第三方服務安全地存取其 Amazon S3 型資料湖中的資料。您可以使用 Lake Formation 做為授權引擎，透過 AWS Glue ETL、Amazon Athena、Amazon EMR 和 Redshift Spectrum 等整合 AWS 服務來管理或強制執行資料湖的許可。Lake Formation 提供兩個整合服務的選項：

1. Lake Formation 應用程式整合設定：Lake Formation 可根據有效許可，將 AWS STS 權杖形式的縮小範圍臨時憑證提供給已註冊的 Amazon S3 位置，以便授權的應用程式可以代表使用者存取資料。
2. 集中強制執行：Lake Formation 查詢 API 操作從 Amazon S3 擷取資料，並根據有效許可篩選結果。與查詢 API 操作整合的引擎或應用程式可以依賴 Lake Formation 來評估呼叫身分的許可，並根據這些許可安全地篩選資料。第三方查詢引擎只會在篩選的資料上查看和操作。

Lake Formation 憑證販賣不會與 spark sql 查詢整合。登入資料販賣僅適用於透過 AWS Glue ETL 程式庫執行的查詢。

主題

- 使用 Lake Formation 應用程式整合

使用 Lake Formation 應用程式整合

Lake Formation 允許第三方服務與 Lake Formation 整合，並使用 GetTemporaryGlueTableCredentials 和 GetTemporaryGluePartitionCredentials 操作，代表其使用者暫時存取 Amazon S3 資料。這可讓第三方服務使用與其他 AWS 分析服務相同的授權和憑證販賣功能。本節說明如何使用這些 API 操作來整合第三方查詢引擎與 Lake Formation。

這些 API 操作預設為停用。有兩種選項可授權 Lake Formation 整合應用程式：

- 設定每次呼叫應用程式整合 API 操作時驗證的 IAM 工作階段標籤

  如需詳細資訊，請參閱啟用第三方查詢引擎呼叫應用程式整合 API 操作的許可。
- 啟用允許外部引擎存取具有完整資料表存取權之 Amazon S3 位置中的資料的選項

  如果使用者具有完整的資料表存取權，此選項可讓查詢引擎和應用程式在沒有 IAM 工作階段標籤的情況下取得登入資料。它提供查詢引擎和應用程式效能優勢，並簡化資料存取。Amazon EC2 上的 Amazon EMR 能夠利用此設定。

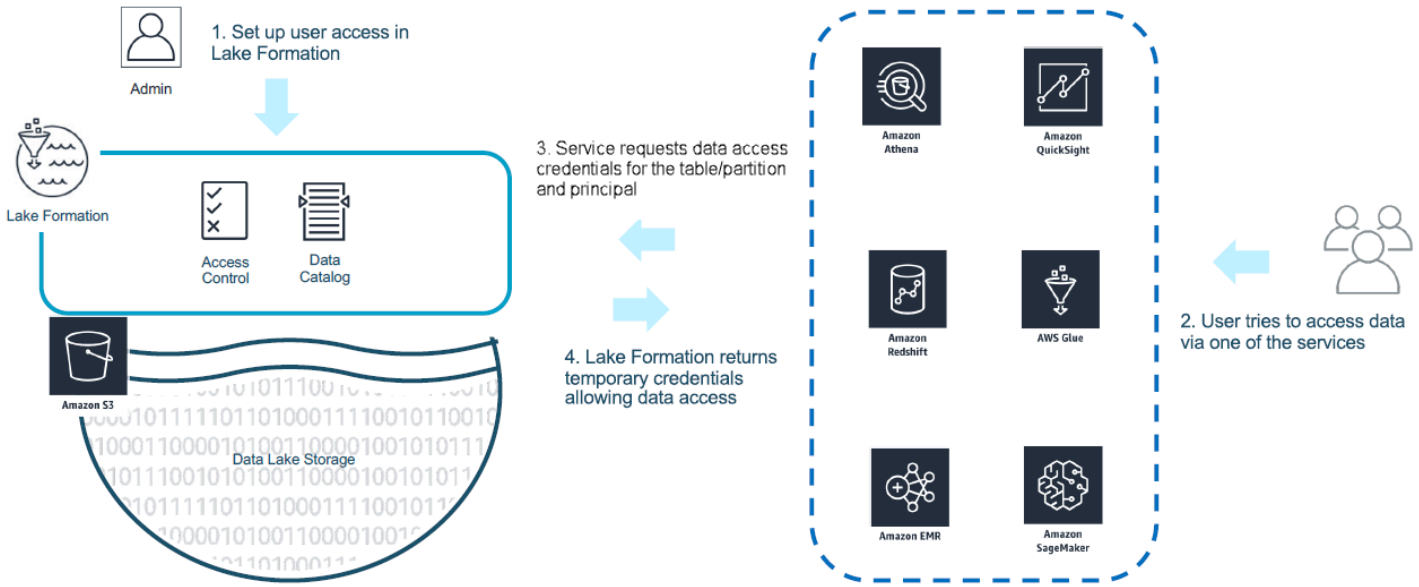
如需詳細資訊，請參閱完整資料表存取的應用程式整合 。

主題

- Lake Formation 應用程式整合的運作方式
- Lake Formation 應用程式整合中的角色和責任
- Lake Formation 應用程式整合 API 操作的工作流程
- 註冊第三方查詢引擎
- 啟用第三方查詢引擎呼叫應用程式整合 API 操作的許可
- 完整資料表存取的應用程式整合

Lake Formation 應用程式整合的運作方式

本節說明如何使用應用程式整合 API 操作，將第三方應用程式 （查詢引擎） 與 整合Lake Formation。



1. Lake Formation 管理員會執行下列活動：
   - 向 Lake Formation 註冊 Amazon S3 位置，方法是提供具有適當許可的 IAM 角色 （用於販賣登入資料），以存取 Amazon S3 位置內的資料
   - 註冊第三方應用程式，以便能夠呼叫 Lake Formation 的憑證販賣 API 操作。請參閱the section called “註冊第三方查詢引擎”
   - 授予使用者存取資料庫和資料表的許可

例如，如果您想要發佈使用者工作階段資料集，其中包含一些包含個人身分識別資訊 (PII) 的資料欄，以限制存取，您可以為這些資料欄指派名為「分類」且值為「敏感」的 LF-TBAC 標籤。接下來，您可以定義許可，允許業務分析師存取使用者工作階段資料，但排除標記為分類 = 敏感的資料欄。

2. 委託人（使用者）向整合服務提交查詢。
3. 整合的應用程式會將請求傳送至 Lake Formation，要求資料表資訊和登入資料來存取資料表。
4. 如果查詢委託人有權存取資料表，Lake Formation 會將登入資料傳回整合式應用程式，以允許資料存取。

> **Note**
>
> 傳送登入資料時，Lake Formation 無法存取基礎資料。

5. 整合的服務會從 Amazon S3 讀取資料、根據收到的政策篩選資料欄，並將結果傳回給委託人。

> **Important**
>
> Lake Formation 憑證販賣 API 操作會啟用分散式強制執行，並明確拒絕失敗（關閉失敗）模型。這引入了客戶、第三方服務和 Lake Formation 之間的第三方安全模型。受信任的整合式服務可正確強制執行Lake Formation許可（分散式強制執行）。

整合服務負責根據從 傳回的政策篩選從 Amazon S3 讀取的資料，Lake Formation然後再將篩選的資料傳回給使用者。整合式服務遵循關閉失敗模型，這表示如果他們無法強制執行必要的Lake Formation許可，則必須讓查詢失敗。

Lake Formation 應用程式整合中的角色和責任

以下是啟用第三方應用程式整合的角色及其相關責任 AWS Lake Formation。

角色	責任
客戶	• 啟用 Lake Formation 應用程式整合設定（請參閱 the section called “註冊第三方查詢引擎”）。 • 向 Lake Formation 明確註冊核准的第三方（請參閱 the section called “註冊第三方查詢引擎”）。

角色	責任
	• 使用 Lake Formation 許可測試和驗證第三方解決方案。 • 監控和稽核 Lake Formation 憑證販賣 API 操作的第三方使用情況。
第三方	• 公開記錄每個軟體修訂的支援功能，並提供正確啟用功能的指示。 • 呼叫 Lake Formation 憑證販賣 API 操作（根據文件）時，準確公告支援的功能。 • 安全地存放和處理提供的登入資料，以避免登入資料洩漏和權限提升。 • 根據支援的功能強制執行許可，並僅將篩選的資料傳回給使用者 • 無法正確強制執行必要許可時，查詢失敗
AWS Lake Formation	• 正確衍生並傳回指定委託人的有效許可。 • 根據 API 操作call-by-call驗證第三方支援的功能。 • 只有在引擎公告的功能符合目錄資源上定義的功能時，才會傳回縮小範圍的 IAM 登入資料，否則會傳回錯誤。

Lake Formation 應用程式整合 API 操作的工作流程

以下是應用程式整合 API 操作的工作流程：

1. 使用者使用整合的第三方查詢引擎提交資料查詢或請求。查詢引擎會擔任代表使用者或使用者群組的 IAM 角色，並擷取呼叫應用程式整合 API 操作時要使用的信任憑證。
2. 查詢引擎會呼叫 GetUnfilteredTableMetadata，如果它是分割資料表，則查詢引擎會呼叫 從 Data Catalog GetUnfilteredPartitionsMetadata擷取中繼資料和政策資訊。
3. Lake Formation 會執行請求的授權。如果使用者沒有資料表的適當許可，則會擲回 AccessDeniedException。
4. 作為請求的一部分，查詢引擎會傳送其支援的篩選。陣列中可以傳送兩個旗標：COLUMN_PERMISSIONS 和 CELL_FILTER_PERMISSION。如果查詢引擎不支援任何這些功能，且該功能的資料表上存在政策，則會擲回 PermissionTypeMismatchException，且查詢會失敗。這是為了避免資料外洩。
5. 傳回的回應包含下列項目：
   - 資料表的整個結構描述，以便查詢引擎可以使用它從儲存體剖析資料。
   - 使用者可存取的授權資料欄清單。如果授權資料欄清單是空的，則表示使用者具有DESCRIBE許可，但沒有SELECT許可，且查詢失敗。

- 標記 `IsRegisteredWithLakeFormation`，指出 Lake Formation 是否可以將登入資料提供給此資源資料。如果傳回 false，則應該使用客戶的登入資料來存取 Amazon S3。
- 如果`CellFilters`有任何 應套用至資料列的 清單。此清單包含要評估每一列的資料欄和表達式。只有在 CELL_FILTER_PERMISSION 作為請求的一部分傳送，且針對呼叫使用者的資料表有資料篩選條件時，才應填入此項目。

6. 擷取中繼資料之後，查詢引擎會呼叫 `GetTemporaryGlueTableCredentials`或 `GetTemporaryGluePartitionCredentials` 以取得 AWS 登入資料，以從 Amazon S3 位置擷取資料。
7. 查詢引擎會從 Amazon S3 讀取相關物件，根據其在步驟 2 中收到的政策篩選資料，並將結果傳回給使用者。

的應用程式整合 API 操作Lake Formation包含用於設定與第三方查詢引擎整合的其他內容。您可以在登入資料販賣 API 操作區段中查看操作詳細資訊。

註冊第三方查詢引擎

在第三方查詢引擎可以使用應用程式整合 API 操作之前，您需要明確啟用查詢引擎代表您呼叫 API 操作的許可。這在幾個步驟中完成：

1. 您需要指定 AWS 帳戶和 IAM 工作階段標籤，這些標籤需要透過 AWS Lake Formation 主控台、AWS CLI 或 API/SDK 呼叫應用程式整合 API 操作的許可。
2. 當第三方查詢引擎在您的帳戶中擔任執行角色時，查詢引擎必須連接向代表第三方引擎的 Lake Formation 註冊的工作階段標籤。 Lake Formation使用此標籤來驗證請求是否來自核准的引擎。如需工作階段標籤的詳細資訊，請參閱《IAM 使用者指南》中的工作階段標籤。
3. 設定第三方查詢引擎執行角色時，您必須在 IAM 政策中擁有下列最低許可集：

```
{
  "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. 在查詢引擎執行角色上設定角色信任政策，以對可連接到此角色的工作階段標籤索引鍵值對進行精細存取控制。在下列範例中，此角色只允許"engine1"連接工作階段標籤金鑰"LakeFormationAuthorizedCaller"和工作階段標籤值，不允許其他工作階段標籤金鑰值對。

```
{
    "Sid": "AllowPassSessionTags",
    "Effect": "Allow",
    "Principal": {
        "AWS": "arn:aws:iam::111122223333:role/query-execution-role"
    },
    "Action": "sts:TagSession",
    "Condition": {
    "StringLike": {
        "aws:RequestTag/LakeFormationAuthorizedCaller": "engine1"        }
    }
}
```

當 LakeFormationAuthorizedCaller呼叫 STS：AssumeRole API 操作來擷取要使用的查詢引擎登入資料時，工作階段標籤必須包含在 AssumeRole 請求中。傳回的暫時登入資料可用於提出Lake Formation應用程式整合 API 請求。

Lake Formation 應用程式整合 API 操作需要呼叫主體做為 IAM 角色。IAM 角色必須包含具有已向 註冊之預定值的工作階段標籤Lake Formation。此標籤允許 Lake Formation 驗證用於呼叫應用程式整合 API 操作的角色是否允許這樣做。

啟用第三方查詢引擎呼叫應用程式整合 API 操作的許可

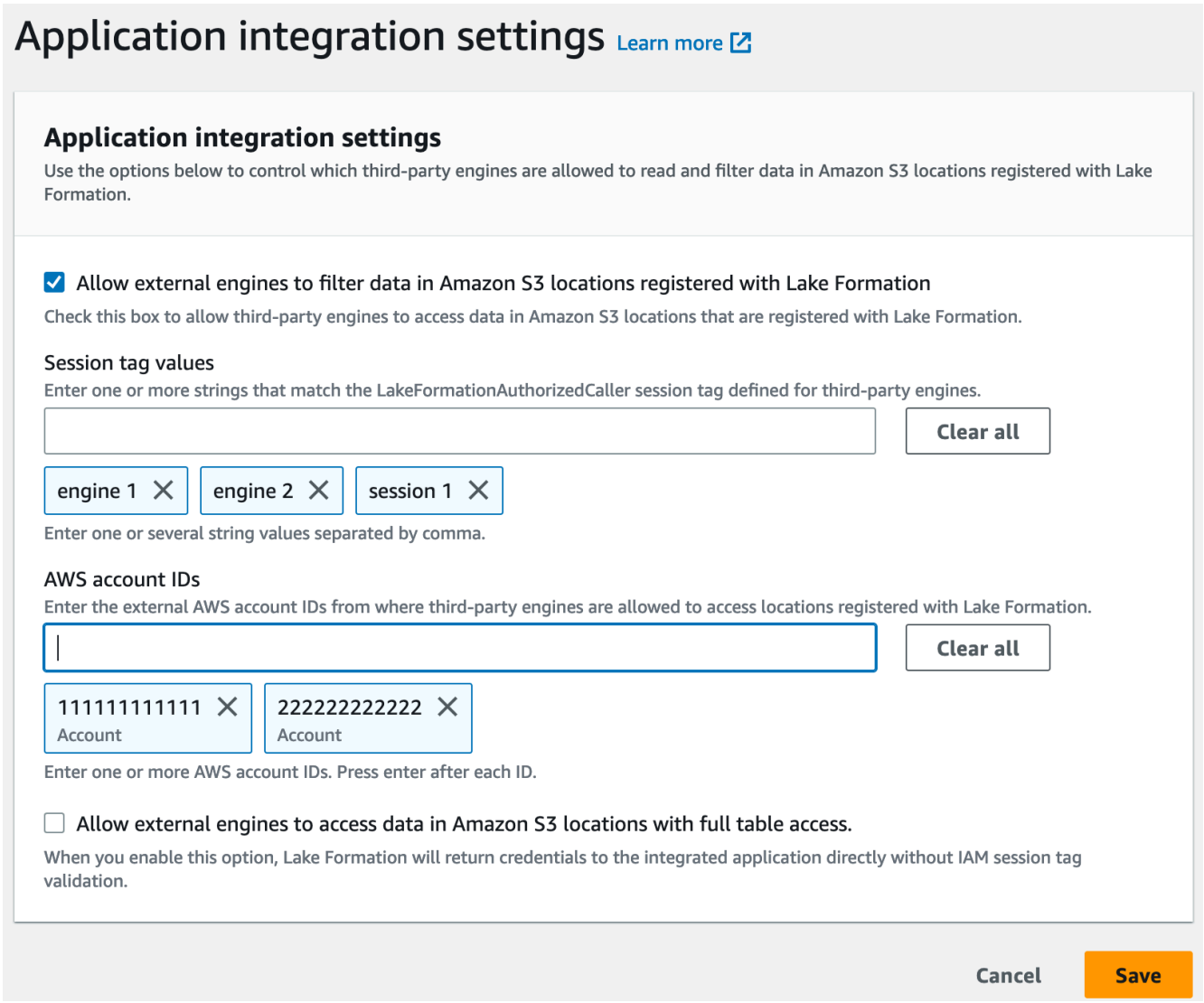
請依照下列步驟，允許第三方查詢引擎透過 主控台、 AWS CLI 或 API/SDK AWS Lake Formation呼叫應用程式整合 API 操作。

Console

若要註冊您的帳戶以進行外部資料篩選：

1. 登入 AWS Management Console，並在 https：//https://console.aws.amazon.com/lakeformation/ 開啟 Lake Formation 主控台。
2. 在左側導覽中，展開管理，然後選擇應用程式整合設定。
3. 在應用程式整合設定頁面上，選擇允許外部引擎篩選向 註冊之 Amazon S3 位置中的資料Lake Formation。
4. 輸入您為第三方引擎建立的工作階段標籤。如需工作階段標籤的資訊，請參閱AWS Identity and Access Management 《 使用者指南》中的在 AWS STS 中傳遞工作階段標籤。
5. 輸入可使用第三方引擎存取未篩選中繼資料資訊的使用者的帳戶 IDs，以及目前帳戶中資源的資料存取憑證。

   您也可以使用 AWS 帳戶 ID 欄位來設定跨帳戶存取。



CLI

使用 `put-data-lake-settings` CLI 命令來設定下列參數。

使用此 AWS CLI 命令時，需要設定三個欄位：

- `allow-external-data-filtering` –（布林值）表示第三方引擎可以存取目前帳戶中資源的未篩選中繼資料資訊和資料存取憑證。
- `external-data-filtering-allow-list` –（陣列）使用第三方引擎時，可存取目前帳戶中資源未篩選中繼資料資訊和資料存取憑證的帳戶 IDs 清單。當 AllowExternalDataFiltering 設為 true 時，ExternalDataFilteringAllowList 屬性必須至少包含一個帳戶 ID。不允許空白清單。

- authorized-sessions-tag-value-list – （陣列） 授權工作階段標籤值的清單 （字串）。如果 IAM 角色登入資料已與授權的鍵/值對連接，則如果工作階段標籤包含在清單中，則會授予工作階段對已設定帳戶中資源上未篩選中繼資料資訊和資料存取登入資料的存取權。授權的工作階段標籤金鑰定義為 *LakeFormationAuthorizedCaller*。
- AllowFullTableExternalDataAccess - （布林值） 當發起人具有完整的資料存取許可時，是否允許第三方查詢引擎取得沒有工作階段標籤的資料存取憑證。

例如：

```
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

使用 PutDataLakeSetting API 操作來設定下列參數。

使用此 API 操作時，需要設定三個欄位：

- AllowExternalDataFiltering – （布林值） 指出第三方引擎是否可以存取目前帳戶中資源的未篩選中繼資料資訊和資料存取憑證。

- ExternalDataFilteringAllowList – （陣列） 可使用第三方引擎存取未篩選中繼資料資訊的帳戶 IDs 清單，以及目前帳戶中資源的資料存取憑證。
- AuthorizedSectionsTagValueList – （陣列） 授權標籤值的清單 （字串）。如果已使用授權標籤連接 IAM 角色登入資料，則工作階段會獲得許可，以存取未篩選的中繼資料資訊和已設定帳戶中資源上的資料存取登入資料。授權的工作階段標籤金鑰定義為 *LakeFormationAuthorizedCaller*。
- AllowFullTableExternalDataAccess - （布林值） 當發起人具有完整的資料存取許可時，是否允許第三方查詢引擎取得沒有工作階段標籤的資料存取憑證。

例如：

```
//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));
}
```

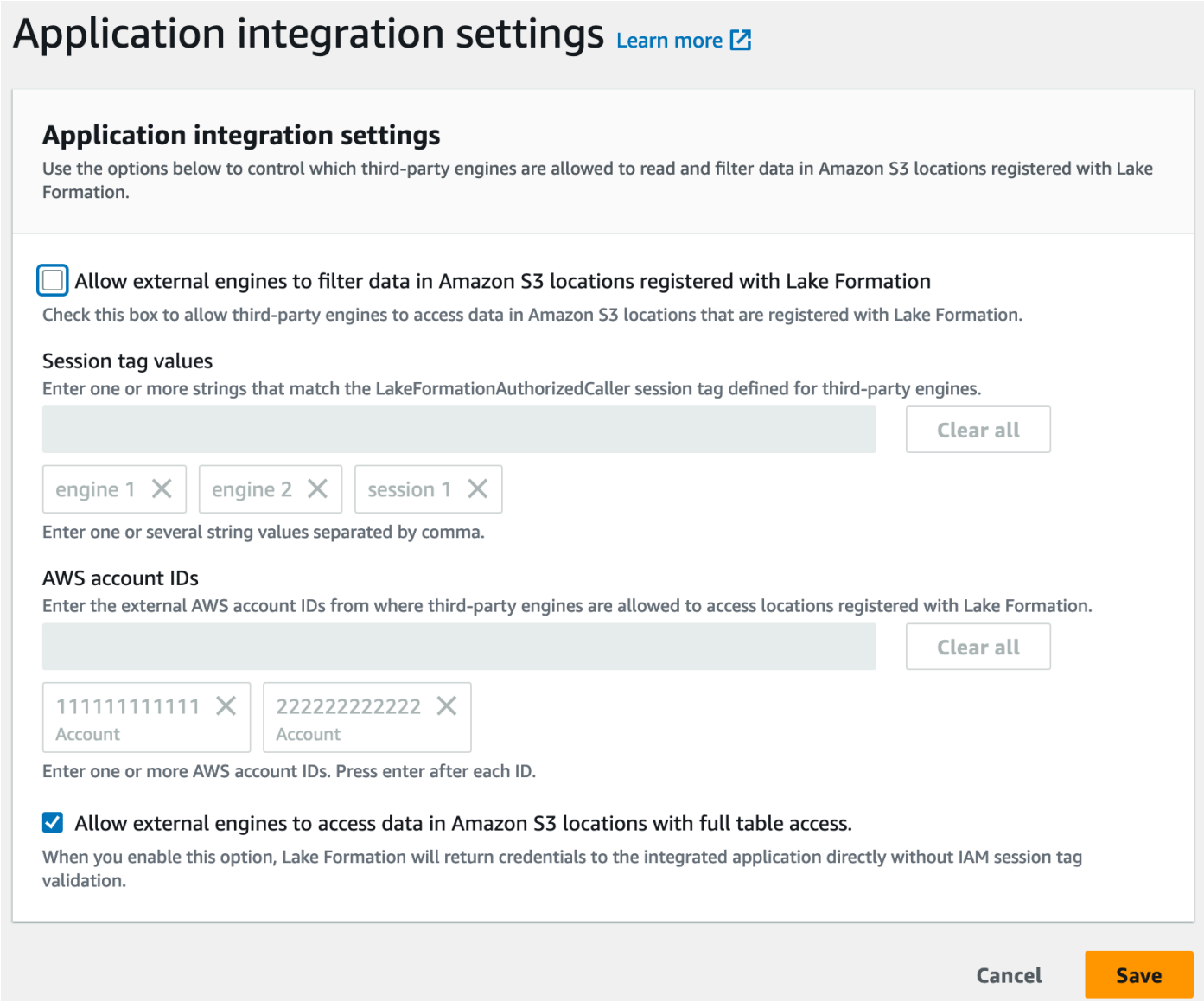
完整資料表存取的應用程式整合

請依照下列步驟，讓第三方查詢引擎在沒有 IAM 工作階段標籤驗證的情況下存取資料：

Console

1. 登入 Lake Formation 主控台，網址為 https：//https://console.aws.amazon.com/lakeformation/。
2. 在左側導覽中，展開管理，然後選擇應用程式整合設定。
3. 在應用程式整合設定頁面上，選擇允許外部引擎使用完整資料表存取選項存取 Amazon S3 位置中的資料。

   當您啟用此選項時，Lake Formation 會直接將登入資料傳回至查詢應用程式，而不需要 IAM 工作階段標籤驗證。



AWS CLI

使用 `put-data-lake-settings` CLI 命令來設定 `AllowFullTableExternalDataAccess` 參數。

```
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
    }
}
```

使用其他 AWS 服務

AWS 服務，例如 Amazon Athena AWS Glue、Amazon Redshift Spectrum 和 Amazon EMR，可用來 AWS Lake Formation 安全地存取向 Lake Formation 註冊的 Amazon S3 位置中的資料。使用 Lake Formation，您可以在 中定義和管理資料表的精細存取控制 (FGAC) 許可 AWS Glue Data Catalog。AWS 這些服務都是 Lake Formation 的可信任發起人，Lake Formation 可透過臨時登入資料存取存放在 Amazon S3 中的資料。如需詳細資訊，請參閱Lake Formation 應用程式整合的運作方式。

為了使用這些功能，Lake Formation 會要求您先註冊 Amazon S3 位置，並將適當的許可指派給 IAM 主體，以存取資料表、資料庫和 Amazon S3 位置。如需詳細資訊，請參閱 管理 Lake Formation 許可。

下表列出 Amazon Athena AWS Glue、Amazon EMR 和 Amazon Redshift Spectrum 支援的 Lake Formation 許可類型，以存取標準 AWS Glue 資料表和交易資料表 (Apache Iceberg、Apache Hudi 和 Linux 基礎 Delta Lake) 中的資料以及 Data Catalog Amazon S3中的資料表中繼資料。

AWSAWS Glue 標準資料表和檢視的 服務和支援的許可類型

AWS 服務	資料表層級許可	資料欄層級許可	資料列和儲存格層級許可
Athena SQL	讀取/寫入存取	讀取存取	讀取存取
Athena Spark	不支援	不支援	不支援
佈建叢集上的 Redshift Spectrum 或 Amazon Redshift 無伺服器	讀取/寫入存取	讀取存取	讀取存取
Amazon EMR (EC2) 上的 Apache Spark	讀取/寫入存取	讀取存取	讀取存取
Amazon EMR (EC2) 上的 Apache Hive	讀取/寫入存取	讀取存取	不支援
EMR Serverless 上的 Apache Spark	讀取/寫入存取	讀取存取	讀取存取

AWS 服務	資料表層級許可	資料欄層級許可	資料列和儲存格層級許可
EMR Serverless 上的 Apache Hive	不支援	不支援	不支援
Amazon EMR on EKS	不支援	不支援	不支援
AWS Glue ETL	讀取/寫入存取	AWS Glue 5.0 或更新版本支援讀取存取。	AWS Glue 5.0 或更新版本支援讀取存取。

考量與限制

- Athena Spark 不支援查詢具有 Lake Formation 許可的資料目錄資料表。
- Athena SAML 型使用者可以透過啟用 SAML 2.0 型聯合來讀取使用 Lake Formation 許可保護的資料來源。SAML 使用者可以將資料插入 Parquet 資料表。
- EMR Serverless 上的 Apache Spark 不支援查詢 Data Catalog 檢視。
- EMR Serverless 上的 Apache Hive 不支援查詢具有 Lake Formation 許可的資料表。
- AWS Glue 5.0 或更高版本支援對資料目錄中由 S3 支援的 Iceberg 和 Hive 資料表進行精細存取控制。此功能可讓您設定 Apache Spark 任務中讀取查詢 AWS Glue 的資料表、資料列、資料欄和儲存格層級存取控制。

  如需詳細資訊，請參閱 AWS Glue 版本。

AWS 交易資料表格式的 服務和支援的許可類型

AWS 服務	Iceberg	Hudi	Delta Lake （原生）	Delta Lake （符號連結資料表）
Athena SQL	支援讀取具有資料表、資料欄、資料列和儲存格層級許可的資料表。寫入操作需要完整資料表存取。	支援對具有資料表、資料欄、資料列和儲存格層級許可的資料表進行讀取和建立操作。不支援寫入操作。	Athena （引擎版本 3) 支援讀取具有資料表、資料欄、資料列和儲存格層級許可的原生 Delta Lake	Athena （引擎版本 3) 支援讀取具有資料表、資料欄、資料列和儲存格層級許可的符號連結 Delta

AWS 服務	Iceberg	Hudi	Delta Lake （原生）	Delta Lake （符號連結資料表）
			資料表。不支援寫入操作。	Lake 資料表。不支援寫入操作。
佈建叢集上的 Redshift Spectrum	支援讀取具有資料表、資料欄、資料列和儲存格層級許可的資料表。不支援寫入操作。	支援讀取具有資料表、資料欄、資料列和儲存格層級許可的資料表。不支援寫入操作。	不支援	支援透過具有資料表、資料欄、資料列和儲存格層級許可的符號連結資訊清單讀取 Delta Lake 資料表。不支援寫入操作。
Amazon EMR (EC2) 上的 Apache Spark	支援讀取具有資料表、資料欄、資料列和儲存格層級許可的資料表。寫入操作需要完整資料表存取。	支援讀取具有資料表、資料欄、資料列和儲存格層級許可的資料表。寫入操作需要完整資料表存取。	支援讀取具有資料表、資料欄、資料列和儲存格層級許可的資料表。不支援寫入操作。	支援讀取具有資料表、資料欄、資料列和儲存格層級許可的資料表。寫入操作需要完整資料表存取。
AWS Glue ETL	AWS Glue 5.0 或更新版本支援讀取具有資料表、資料欄、資料列和儲存格層級許可的資料表。	支援對具有資料表層級許可的資料表進行讀取/寫入。	支援對具有資料表層級許可的資料表進行讀取/寫入。	支援對具有資料表層級許可的資料表進行讀取/寫入。

主題

- AWS Lake Formation 搭配 Amazon Athena 使用
- AWS Lake Formation 搭配 Amazon Redshift Spectrum 使用
- AWS Lake Formation 搭配 使用 AWS Glue
- AWS Lake Formation 搭配 Amazon EMR 使用
- AWS Lake Formation 搭配 QuickSight 使用

- AWS Lake Formation 搭配 AWS CloudTrail Lake 使用

AWS Lake Formation 搭配 Amazon Athena 使用

Amazon Athena 是一種無伺服器查詢服務，可協助您分析存放在 Amazon S3 中的結構化、半結構化和非結構化資料。您可以使用 Athena SQL 查詢 CSV、JSON、Parquet 和 Avro 資料格式的資料。Athena SQL 也支援 Apache Hive、Apache Hudi 和 Apache Iceberg 等資料表格式。Athena 與整合 AWS Glue Data Catalog ，將資料集的中繼資料存放在 Amazon S3 中。Athena 可以使用 Lake Formation 來定義和維護這些資料集的存取控制政策。

以下是您可以搭配 Athena 使用 Lake Formation 的一些常見使用案例。

- 使用 Lake Formation 許可從 Athena 存取 Data Catalog 資源 （資料庫和資料表）。您可以使用具名資源方法或 LF 標籤來定義資料庫和資料表的許可。如需詳細資訊，請參閱：
  - 使用具名資源方法授予資料庫許可
  - Lake Formation 標籤型存取控制

> **Note**
>
> Lake Formation 許可僅適用於使用 Athena SQL 查詢來自 Amazon S3 的來源資料和 Data Catalog 中的中繼資料。
> Athena Spark 不支援查詢具有 Lake Formation 許可的資料目錄資料表。Lake Formation 許可支援資料庫和資料表上的讀取和寫入操作。

> **Note**
>
> 當您使用 LF 標籤來管理 Data Catalog 資源的許可時，無法套用資料篩選條件。

- 透過授予資料欄、資料列和儲存格層級的許可，使用 Lake Formation 中的資料篩選條件來保護 Amazon S3 資料湖中的資料表，以控制查詢結果。請參閱《Amazon Athena 使用者指南》中的分割區投影限制。
- 執行聯合查詢時，對 SAML 型 Athena 使用者可用的資料強制執行精細存取控制。

  Athena JDBC 和 ODBC 驅動程式支援使用 SAML 型身分提供者 (IdP) 設定對資料來源的聯合存取。使用與 Lake Formation 整合的 QuickSight 與您現有的 IAM 角色或 SAML 使用者或群組，以視覺化方式呈現 Athena 查詢結果。

> **Note**
>
> 只有當您使用 JDBC 或 ODBC 驅動程式向 Athena 提交查詢時，SAML 使用者和群組的 Lake Formation 許可才會套用。

如需詳細資訊，請參閱使用 Lake Formation 和 Athena JDBC 和 ODBC 驅動程式來聯合存取 Athena。

> **Note**
>
> 目前，下列區域不支援授權存取 Lake Formation 中的 SAML 身分：
>
> - 中東 (巴林) – me-south-1
> - 亞太區域 (香港) – ap-east-1
> - 非洲 (開普敦) – af-south-1
> - 中國 (寧夏) – cn-northwest-1
> - 亞太區域 (大阪) - ap-northeast-3

- 使用 Lake Formation 中的跨帳戶資料共用 查詢另一個帳戶中的資料表。

> **Note**
>
> 如需對 使用 Lake Formation 許可時限制的詳細資訊`Views`，請參閱考量和限制。

支援交易資料表格式

套用 Lake Formation 許可可讓您保護 Amazon S3 型資料湖中的交易資料。下表列出 Athena 和 Lake Formation 許可中支援的交易資料表格式。Lake Formation 會在 Athena 使用者執行其查詢時強制執行這些許可。

資料表格式	描述和允許的操作	Athena 中支援的 Lake Formation 許可
Apache Hudi	用於簡化增量資料處理和資料管道開發的格式。 Athena 支援在 Amazon S3 資料集上使用 Apache Hudi 資料表格式建立和讀取操作，適用於寫入時複製 (CoW) 和讀取時合併 (MoR) Hudi 資料表類型。Athena 不支援 Hudi 資料表上的寫入操作。 使用 Athena 查詢 Hudi 資料集。	使用 Lake Formation 中的資料篩選和儲存格層級安全性 使用資料表、資料欄、資料列和儲存格層級許可來保護 Hudi 資料表。
Apache Iceberg	一種開放的資料表格式，可管理大型檔案集合做為資料表，並支援現代分析資料湖操作，例如記錄層級插入、更新、刪除和時間歷程查詢。 如需 Athena 支援 Iceberg 資料表的詳細資訊，請參閱使用 Iceberg 資料表。	支援資料表、資料欄、資料列和儲存格層級許可。目前，Lake Formation 不支援管理 Open Table Formats 中 `VACUUM`、 `MERGEUPDATE`和等寫入操作`OPTIMIZE`的許可。
Linux Foundation Delta Lake	Delta Lake 是一項開放原始碼專案，可協助實作常見於 Amazon S3 或 Hadoop 分散式檔案系統 (HDFS) 的現代資料湖架構。 Athena 支援 AWS Glue Data Catalog 從 Delta Lake 資料表使用以符號連結為基礎的資訊清單資料表定義在 上建立的 Delta lake 資料表。	symlink 資料表和原生 Delta Lake 資料表支援資料表、資料欄、資料列和儲存格層級許可。

資料表格式	描述和允許的操作	Athena 中支援的 Lake Formation 許可
	如需詳細資訊，請參閱使用爬蟲程式爬 AWS Glue 取 Delta Lake 資料表。 Athena （引擎版本 3) 支援讀取原生 Delta Lake 資料表。 如需詳細資訊，請參閱使用 AWS Glue 爬蟲程式介紹原生 Delta Lake 資料表支援。	

其他資源

部落格文章、影片和研討會

- 使用 Amazon Athena 查詢 Amazon S3 資料湖中的 Apache Hudi 資料集 Amazon Athena
- 使用 Amazon Athena、Amazon EMR 和 建置 Apache Iceberg 資料湖 AWS Glue
- 使用 Athena 和 Apache Iceberg 在 Amazon S3 上插入、更新、刪除
- 以 LF 標籤為基礎的存取控制Lake Formation 關於查詢資料湖的研討會。

AWS Lake Formation 搭配 Amazon Redshift Spectrum 使用

Amazon Redshift Spectrum 可讓您查詢和擷取 Amazon S3 資料湖中的資料，而無需將資料載入 Amazon Redshift 叢集節點。

Redshift Spectrum 支援兩種方式來註冊已啟用 Lake Formation 的外部 AWS Glue 資料目錄。

- 使用具有 Data Catalog 許可的叢集連接 IAM 角色

  若要建立 IAM 角色，請遵循下列程序中概述的步驟。

  控制對 的存取 AWS Glue Data Catalog
- 使用設定為管理外部 AWS Glue Data Catalog 資源存取的聯合 IAM 身分

Redshift Spectrum 支援使用聯合 IAM 身分查詢 Lake Formation 資料表。IAM 身分可以是 IAM 使用者或 IAM 角色。如需 Redshift Spectrum 中 IAM 聯合身分的詳細資訊，請參閱使用聯合身分管理對本機資源和 Redshift Spectrum 外部資料表的 Amazon Redshift 存取。

透過 Lake Formation 與 Redshift Spectrum 整合，您可以在向 Lake Formation 註冊資料後定義資料表上的資料列、資料欄和儲存格層級存取控制許可。

如需詳細資訊，請參閱搭配 使用 Redshift Spectrum AWS Lake Formation。

Redshift Spectrum 支援 Lake Formation 受管外部結構描述資料表上的讀取或SELECT查詢。

如需詳細資訊，請參閱建立 Redshift Spectrum 的外部結構描述。

支援交易資料表類型

此資料表列出 Redshift Spectrum 中支援的交易資料表格式和適用的 Lake Formation 許可。

支援的資料表格式

資料表格式	描述和允許的操作	Redshift Spectrum 中支援的 Lake Formation 許可
Apache Hudi	用於簡化增量資料處理和資料管道開發的格式。 Redshift Spectrum 支援在 Amazon S3 上使用 Apache Hudi 寫入時複製 (CoW) 資料表格式的插入、刪除和 upsert 寫入操作。 如需詳細資訊，請參閱為 Apache Hudi 中管理的資料建立外部資料表。	使用 Lake Formation 中的資料篩選和儲存格層級安全性 保護使用資料表、資料欄、資料列和儲存格層級許可的 Hudi 資料表。
Apache Iceberg	一種開放的資料表格式，可管理大型檔案集合做為資料表，並支援現代分析資料湖操作，例如記錄層級插入、更新、刪除和時間歷程查詢。	Redshift Spectrum 支援 Apache Iceberg 資料表進行查詢。

資料表格式	描述和允許的操作	Redshift Spectrum 中支援的 Lake Formation 許可
	如需詳細資訊，請參閱搭配 Amazon Redshift 使用 Apache Iceberg 資料表。	
Linux Foundation Delta Lake	Delta Lake 是一項開放原始碼專案，可協助實作常見於 Amazon S3 或 Hadoop 分散式檔案系統 (HDFS) 的現代資料湖架構。 Redshift Spectrum 支援查詢 Delta Lake 資料表。如需詳細資訊，請參閱為 Delta Lake 中管理的資料建立外部資料表。	支援資料表、資料欄、資料列和儲存格層級許可。

其他資源

部落格文章和研討會

- 使用 集中管理資料湖， AWS Lake Formation 同時使用 Amazon Redshift Spectrum 啟用現代資料架構
- 使用 Redshift Spectrum 查詢 Amazon S3 資料湖中的 Apache HUDI 寫入時複製 (CoW) 資料表

AWS Lake Formation 搭配 使用 AWS Glue

資料工程師和 DevOps 專業人員使用 AWS Glue 搭配擷取、轉換和載入 (ETL) 搭配 Apache Spark，在 Amazon S3 中的資料集上執行轉換，並將轉換的資料載入資料湖和資料倉儲，以進行分析、機器學習和應用程式開發。對於在 Amazon S3 中存取相同資料集的不同團隊，必須根據其角色授予和限制許可。

AWS Lake Formation 以 為基礎 AWS Glue，服務會以下列方式互動：

- Lake Formation 和AWS Glue共用相同的資料目錄。
- 下列 Lake Formation 主控台功能會叫用AWS Glue主控台：

  - 任務 – 如需詳細資訊，請參閱《 AWS Glue 開發人員指南》中的新增任務。
  - 爬蟲程式 – 如需詳細資訊，請參閱《 AWS Glue 開發人員指南》中的使用爬蟲程式編目資料表。
- 使用 Lake Formation 藍圖時產生的工作流程是AWS Glue工作流程。您可以在 Lake Formation 主控台和AWS Glue主控台中檢視和管理這些工作流程。
- Lake Formation 提供機器學習轉換，並建置在 AWS Glue API 操作上。您可以在 AWS Glue主控台上建立和管理機器學習轉換。如需詳細資訊，請參閱《 AWS Glue 開發人員指南》中的Machine Learning轉換。

您可以使用 Lake Formation 精細存取控制來管理現有的 Data Catalog 資源和 Amazon S3 資料位置。

> **Note**
>
> AWS Glue 5.0 或更高版本支援對由 S3 支援的 Iceberg 和 Hive 資料表進行精細存取控制。此功能可讓您設定 Apache Spark 任務中讀取查詢 AWS Glue 的資料表、資料列、資料欄和儲存格層級存取控制。

支援交易資料表類型

套用 Lake Formation 許可可讓您保護 Amazon S3 型資料湖中的交易資料。下表列出 AWS Glue 和 Lake Formation 許可中支援的交易資料表格式。Lake Formation 會強制執行這些 AWS Glue 操作許可。

支援的資料表格式

資料表格式	描述和允許的操作	中支援的 Lake Formation 許可 AWS Glue
Apache Hudi	用於簡化增量資料處理和資料管道開發的開放資料表格式。 如需範例，請參閱在 中使用 Hudi 架構 AWS Glue。	資料表層級許可可用於 Hudi 資料表。 如需詳細資訊，請參閱限制。
Apache Iceberg	以資料表形式管理大型檔案集合的開放資料表格式。	AWS Glue 5.0 版和更新版本可讓您設定 Iceberg 資料表 AWS Glue Apache Spark 任務內讀

資料表格式	描述和允許的操作	中支援的 Lake Formation 許可 AWS Glue
	如需範例，請參閱在 中使用 Iceberg 架構 AWS Glue。	取查詢的資料表、資料列、資料欄和儲存格層級存取控制。 如需詳細資訊，請參閱限制。
Linux Foundation Delta Lake	Delta Lake 是一項開放原始碼專案，可協助實作常見於 Amazon S3 或 Hadoop 分散式檔案系統 (HDFS) 的現代資料湖架構。 如需範例，請參閱在 中使用 Delta Lake 架構 AWS Glue。	資料表層級許可可用於 Delta Lake 資料表。 如需詳細資訊，請參閱限制。

其他資源

部落格文章和儲存庫

- 使用 AWS Glue 連接器透過 ACID 交易讀取和寫入 Apache Iceberg 資料表，並執行時間歷程
- 使用 AWS Glue 自訂連接器寫入 Apache Hudi 資料表
- AWS Cloudformation 範本和 pyspark 程式碼範例的儲存庫，用於使用 AWS Glue、Apache Hudi 和 Amazon S3 分析串流資料。

AWS Lake Formation 搭配 Amazon EMR 使用

Amazon EMR 是一種靈活的 AWS 受管叢集平台，您可以在支援的大數據架構上執行任何自訂程式碼，例如 Hadoop Map-Reduce、Spark、Hive、Presto 等。組織也會使用 Amazon EMR 跨高度分散式叢集執行批次和串流資料處理應用程式。在 Amazon EMR 上使用 Apache Spark，您可以在其許可由 Lake Formation 管理的資料庫和資料表上執行資料轉換和自訂程式碼。

部署 Amazon EMR 有三個選項：

- EC2 上的 EMR
- EMR Serverless

- Amazon EMR on EKS

如需詳細資訊，請參閱整合 Amazon EMR 與 Lake Formation 或使用 EMR Serverless 與 AWS Lake Formation 進行精細存取控制

支援交易資料表格式

當您使用 Spark SQL 讀取和寫入資料時，Amazon EMR 6.15.0 版和更新版本包括對 Apache Hudi、Apache Iceberg 和 Delta Lake 資料表格式的 Lake Formation 資料表、資料列、資料欄和儲存格層級存取控制許可的支援。

如需限制，請參閱使用 Lake Formation 的 Amazon EMR 考量事項。

支援的資料表格式

資料表格式	描述和允許的操作	Amazon EMR 中支援的 Lake Formation 許可
Apache Hudi	用於簡化增量資料處理和資料管道開發的開放資料表格式。 如需支援的操作清單，請參閱 Apache Hudi 和 Lake Formation。	Amazon EMR 使用 Apache Hudi 來支援資料表、資料列、資料欄和儲存格層級存取控制。
Apache Iceberg	以資料表形式管理大型檔案集合的開放資料表格式。 如需支援的操作清單，請參閱 Apache Iceberg 和 Lake Formation。	Amazon EMR 使用 Apache Iceberg 來支援資料表、資料列、資料欄和儲存格層級存取控制。
Linux Foundation Delta Lake	Delta Lake 是一項開放原始碼專案，可協助實作常見於 Amazon S3 或 Hadoop 分散式檔案系統 (HDFS) 的現代資料湖架構。	Amazon EMR 支援使用 Delta Lake 資料表進行資料表、資料列、資料欄和儲存格層級存取控制。

資料表格式	描述和允許的操作	Amazon EMR 中支援的 Lake Formation 許可
	如需支援的操作清單，請參閱 Delta Lake 和 Lake Formation 。	

其他資源

使用者指南、部落格文章和研討會

- 使用執行期角色與 Amazon EMR 整合
- 使用 Amazon EMR on EKS 快速開始使用 Apache Hudi、Apache Iceberg 和 Delta Lake
- 搭配 EMR Serverless 使用 Delta Lake OSS

AWS Lake Formation 搭配 QuickSight 使用

QuickSight 支援探索使用 Athena 在 Amazon S3 中由 Lake Formation 許可管理的資料集。

QuickSight 的標準版和企業版使用者都與 Lake Formation 整合，但略有不同。

- 企業版 – 將精細存取控制 (FGAC) 許可授予個別 QuickSight 使用者和群組，以存取資料庫和資料表。
- 標準版本 – 授予 IAM 角色存取資料庫和資料表的許可。

> **Note**
>
> 根據預設，QuickSight 會使用名為 的角色`aws-quicksight-service-role-v0`。您也可以定義具有必要許可的自訂角色，讓 QuickSight 能夠存取 Athena。

如需詳細資訊，請參閱透過 授權連線 AWS Lake Formation

其他資源

部落格文章

- 在 中啟用 QuickSight 作者的精細許可 AWS Lake Formation
- 使用 AWS Lake Formation 和 QuickSight 安全地分析您的資料

AWS Lake Formation 搭配 AWS CloudTrail Lake 使用

AWS CloudTrail Lake 支援使用 來探索事件資料存放區 Amazon Athena ，其中包含精細的許可 AWS Lake Formation。

> **Note**
>
> CloudTrail Lake 只能透過 查詢 Amazon Athena。

若要向 Lake Formation 註冊 CloudTrail Lake 事件資料存放區，請參閱聯合事件資料存放區。

使用 記錄 AWS Lake Formation API 呼叫 AWS CloudTrail

AWS Lake Formation 已與 服務整合 AWS CloudTrail，此服務提供由使用者、角色或 Lake Formation 中的 AWS 服務所採取之動作的記錄。CloudTrail 會將所有 Lake Formation API 呼叫擷取為事件。擷取的呼叫包括來自 Lake Formation 主控台的呼叫 AWS Command Line Interface、 和對 Lake Formation API 動作的程式碼呼叫。如果您建立線索，您可以啟用 CloudTrail 事件持續交付至 Amazon S3 儲存貯體，包括 Lake Formation 的事件。即使您未設定追蹤，依然可以透過 CloudTrail 主控台中的事件歷史記錄檢視最新事件。您可以使用 CloudTrail 所收集的資訊，判斷對 Lake Formation 提出的請求、提出請求的 IP 地址、提出請求的人員、提出請求的時間，以及其他詳細資訊。

若要進一步了解 CloudTrail，請參閱「AWS CloudTrail 使用者指南」。

CloudTrail 中的 Lake Formation 資訊

當您建立新 AWS 帳戶時，CloudTrail 預設為啟用。在 Lake Formation 中發生活動時，該活動會記錄為 CloudTrail 事件，以及事件歷史記錄中的其他 AWS 服務事件。事件即為來自任何來源的單一請求，其中包含請求動作、動作日期和時間，以及請求參數的相關資訊。此外，每個事件或日誌項目都包含產生請求的人員相關資訊。身分資訊可協助您判斷下列事項：

- 請求是使用根或 AWS Identity and Access Management (IAM) 使用者登入資料提出。
- 提出該請求時，是否使用了特定角色或聯合身分使用者的暫時安全憑證。
- 請求是否由其他 AWS 服務提出。

如需詳細資訊，請參閱 CloudTrail userIdentity 元素。

您可以檢視、搜尋和下載您 AWS 帳戶的最近事件。如需詳細資訊，請參閱「使用 CloudTrail 事件歷史記錄檢視事件」。

若要持續記錄您 AWS 帳戶中的事件，包括 Lake Formation 的事件，請建立追蹤。線索能讓 CloudTrail 將日誌檔案交付至 Amazon S3 儲存貯體。根據預設，在主控台建立線索時，該線索會套用到所有 AWS 區域。線索會記錄 AWS 分割區中所有區域的事件，並將日誌檔案交付至您指定的 Amazon S3 儲存貯體。此外，您可以設定其他服務 AWS Amazon Athena，例如，進一步分析和處理 CloudTrail 日誌中收集的事件資料。CloudTrail 也可以將日誌檔案交付至 Amazon CloudWatch Logs 和 CloudWatch Events。

如需詳細資訊，請參閱下列內容：

- 建立追蹤的概觀

- CloudTrail 支援的服務和整合
- 設定 CloudTrail 的 Amazon SNS 通知
- 接收多個區域的 CloudTrail 日誌檔案和接收多個帳戶的 CloudTrail 日誌檔案

了解 Lake Formation 事件

CloudTrail 會記錄所有 Lake Formation API 動作，並記錄在 AWS Lake Formation 開發人員指南中。例如，對 PutDataLakeSettings、GrantPermissions 以及 RevokePermissions 動作發出的呼叫會在 CloudTrail 日誌檔案中產生項目。

下列範例顯示 GrantPermissions動作的 CloudTrail 事件。項目包括授予許可的使用者 (datalake_admin)、授予許可的委託人 (datalake_user1)，以及授予許可的 ()CREATE_TABLE。項目也會顯示授予失敗，因為未在resource引數中指定目標資料庫。

```
{
    "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"
}
```

下一個範例顯示 GetDataAccess動作的 CloudTrail 日誌項目。主體不會直接呼叫此 API。相反地，每當委託人或整合 AWS 服務請求暫時登入資料，以存取向 Lake Formation 註冊的資料湖位置中的資料時，GetDataAccess就會記錄 。

```
{
    "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"
    },
...
}
```

另請參閱

- 跨帳戶 CloudTrail 記錄

Lake Formation 最佳實務、考量和限制

使用本節快速尋找其中的最佳實務、考量事項和限制 AWS Lake Formation。

如需 的服務資源或操作數量上限，請參閱服務配額 AWS 帳戶。

主題

- 跨帳戶資料共用最佳實務和考量事項
- 服務連結角色限制
- 跨區域資料存取限制
- Data Catalog 檢視的考量和限制
- 資料篩選限制
- 混合存取模式的考量和限制
- 將 Amazon Redshift 資料倉儲資料帶入 的限制 AWS Glue Data Catalog
- S3 資料表目錄整合限制
- Hive 中繼資料存放區資料共用的考量和限制
- Amazon Redshift 資料共用限制
- IAM Identity Center 整合限制
- Lake Formation 標籤型存取控制最佳實務和考量事項
- 屬性型存取控制考量事項、限制和支援的區域

跨帳戶資料共用最佳實務和考量事項

Lake Formation 跨帳戶功能可讓使用者安全地跨多個 AWS 組織共用分散式資料湖 AWS 帳戶，或直接與另一個帳戶中的 IAM 主體共用，提供對 Data Catalog 中繼資料和基礎資料的精細存取。

使用 Lake Formation 跨帳戶資料共用時，請考慮下列最佳實務：

- 您可以對自己 AWS 帳戶中的委託人進行 Lake Formation 許可授予的數量沒有限制。不過，Lake Formation 會使用 AWS Resource Access Manager (AWS RAM) 容量進行跨帳戶授予，您的帳戶可以使用具名資源方法進行授予。若要最大化 AWS RAM 容量，請遵循下列具名資源方法的最佳實務：
  - 使用新的跨帳戶授予模式（跨帳戶版本設定下的第 3 版及更高版本）與外部 共用資源 AWS 帳戶。 如需詳細資訊，請參閱更新跨帳戶資料共用版本設定。

- 將 AWS 帳戶安排在組織中，並將許可授予組織或組織單位。對組織或組織單位的授予會計為一個授予。

  授予組織或組織單位也不需要接受授予的 AWS Resource Access Manager (AWS RAM) 資源共享邀請。如需詳細資訊，請參閱存取和檢視共用的 Data Catalog 資料表和資料庫。

- 使用特殊的所有資料表萬用字元來授予資料庫中所有資料表的許可，而不是授予資料庫中許多個別資料表的許可。授予所有資料表會計為單一授予。如需詳細資訊，請參閱授予 Data Catalog 資源的許可。

> **Note**
>
> 如需請求更高限制的資源共用數量的詳細資訊 AWS RAM，請參閱 中的AWS 服務配額AWS 一般參考。

- 您必須建立共用資料庫的資源連結，該資料庫才會出現在 Amazon Athena 和 Amazon Redshift Spectrum 查詢編輯器中。同樣地，若要能夠使用 Athena 和 Redshift Spectrum 查詢共用資料表，您必須建立資料表的資源連結。資源連結接著會出現在查詢編輯器的資料表清單中。

  您可以使用所有資料表萬用字元來授予資料庫中所有資料表的許可，而不是為許多個別資料表建立資源連結以進行查詢。然後，當您為該資料庫建立資源連結並在查詢編輯器中選取該資料庫資源連結時，您可以存取該資料庫中查詢的所有資料表。如需詳細資訊，請參閱建立資源連結。

- 當您直接與另一個帳戶中的主體共用資源時，收件人帳戶中的 IAM 主體可能沒有建立資源連結的許可，能夠使用 Athena 和 Amazon Redshift Spectrum 查詢共用資料表。資料湖管理員可以建立預留位置資料庫，並將`CREATE_TABLE`許可授予`ALLIAMPrincipal`群組，而不是為每個共用的資料表建立資源連結。然後，收件人帳戶中的所有 IAM 主體都可以在預留位置資料庫中建立資源連結，並開始查詢共用資料表。

  請參閱 `ALLIAMPrincipals`中授予許可給 的範例 CLI 命令使用具名資源方法授予資料庫許可。

- 當跨帳戶許可直接授予委託人時，只有授予的收件人可以檢視這些許可。收件人 AWS 帳戶中的資料湖管理員無法檢視這些直接授權。
- Athena 和 Redshift Spectrum 支援資料欄層級存取控制，但僅用於包含，而非排除。AWS Glue ETL 任務不支援資料欄層級存取控制。
- 與 AWS 您的帳戶共用資源時，您只能將資源的許可授予帳戶中的使用者。您無法將資源的許可授予其他 AWS 帳戶、組織（甚至您自己的組織）或`IAMAllowedPrincipals`群組。
- 您無法將資料庫`Super`上的 `DROP`或 授予外部帳戶。

- 在您刪除資料庫或資料表之前，撤銷跨帳戶許可。否則，您必須刪除其中孤立的資源共用 AWS Resource Access Manager。

> 另請參閱
>
> - Lake Formation 標籤型存取控制最佳實務和考量事項
> - CREATE_TABLE 中的 Lake Formation 許可參考，以取得更多跨帳戶存取規則和限制。

服務連結角色限制

服務連結角色是直接連結的特殊 IAM 角色類型 AWS Lake Formation。此角色具有預先定義的許可，允許 Lake Formation 代表您跨 AWS 服務執行動作。

使用服務連結角色 (SLR) 向 Lake Formation 註冊資料位置時，適用下列限制。

- 一旦建立，您就無法修改服務連結角色政策。
- 服務連結角色不支援跨帳戶共用加密的目錄資源。加密的資源需要特定的 AWS KMS 金鑰許可。服務連結角色具有預先定義的許可，不包括跨帳戶使用加密目錄資源的能力。
- 註冊多個 Amazon S3 位置時，使用服務連結角色可能會導致您快速超過 IAM 政策限制。這是因為使用服務連結角色時， 會為您 AWS 撰寫政策，並遞增為一個包含所有註冊的大型區塊。您可以更有效率地撰寫客戶受管政策、跨多個政策分配許可，或對不同的區域使用不同的角色。
- EC2 上的 Amazon EMR 無法存取您使用服務連結角色註冊資料位置的資料。
- 服務連結角色操作會略過您的 AWS 服務控制政策。
- 當您使用服務連結角色註冊資料位置時，它會更新最終一致性的 IAM 政策。如需詳細資訊，請參閱《IAM 使用者指南》中的疑難排解 IAM 文件。

跨區域資料存取限制

Lake Formation 支援跨 查詢 Data Catalog 資料表 AWS 區域。您可以使用 Amazon Athena、Amazon EMR 和 AWS Glue ETL 從其他區域存取區域中的資料，方法是在其他區域中建立指向來源資料庫和資料表的資源連結。透過跨區域資料表存取，您可以跨區域存取資料，而無需將基礎資料或中繼資料複製到 Data Catalog。

下列限制適用於跨區域資料表存取。

- Lake Formation 不支援使用 Amazon Redshift Spectrum 從其他區域查詢 Data Catalog 資料表。
- 在 Lake Formation 主控台中，資料庫和資料表檢視不會顯示來源區域資料庫/資料表名稱。
- 若要從另一個區域檢視共用資料庫下的資料表清單，您必須先建立共用資料庫的資源連結，然後選取資源連結，然後選擇檢視資料表。
- Lake Formation 不支援 SAML 使用者發出的跨區域資源連結呼叫。
- Lake Formation 的跨區域功能不涉及資料傳輸的額外費用。

Data Catalog 檢視的考量和限制

下列考量和限制適用於 Data Catalog 檢視。

- 您無法從 Lake Formation 主控台建立 Data Catalog 檢視。您可以使用 AWS CLI 或 SDK 建立檢視。
- 您可以從 10 個資料表建立 Data Catalog 檢視。這是硬性限制。檢視的基礎參考資料表可以屬於相同 AWS 帳戶中的相同資料庫或不同資料庫。
- 如需使用 Redshift 建立 Data Catalog 檢視的其他考量和限制，請參閱《Amazon Redshift 資料庫開發人員指南》中的 Data Catalog 檢視考量和限制一節。對於 Athena，請參閱《Amazon Athena 使用者指南》中的 Data Catalog 檢視考量和限制一節。
- 您可以在混合存取模式和 Lake Formation 模式中，在向 Lake Formation 註冊的資料表上建立 Data Catalog 檢視。

  搭配 Lake Formation 混合存取模式使用 Data Catalog 檢視時，建議確保使用檢視的主體選擇加入檢視中參考之基礎資料表的 Lake Formation 許可，而不授予存取權。這可確保基底資料表不會透過 IAM AWS Glue 許可向消費者公開。
- 跨帳戶共用版本沒有共用檢視的限制。
- 當您針對已建立的檢視方言使用 `ALTER VIEW`陳述式時，檢視的版本化方式與 Data Catalog 資料表相同。您無法復原至先前的檢視，因為檢視版本會隨著基礎資料變更而變更。您可以刪除檢視版本，它將預設為下一個可用的最新版本。當您變更檢視版本時，請確定您的資料與選取的檢視版本結構描述同步。
- 不會引入新的 Data Catalog APIs。現有的 `UpdateTable`、 `CreateTable``DeleteTable`和 `GetTable` APIs會更新。
- Amazon Redshift 一律會從具有字串的資料表中使用 varchar 資料欄建立檢視。從其他引擎新增方言時，您必須將字串資料欄轉換為明確長度的 varchar。

- 將資料湖許可授予資料庫中All tables的 ，會導致承授者具有資料庫中所有資料表和檢視的許可。
- 您無法建立檢視：
  - 該參考其他檢視。
  - 當參考資料表是資源連結時。
  - 當參考資料表位於另一個 帳戶中時。
  - 從外部 Hive 中繼存放區。

資料篩選限制

當您授予資料目錄資料表的 Lake Formation 許可時，您可以包含資料篩選規格，以限制對與 Lake Formation 整合之查詢結果和引擎中特定資料的存取。Lake Formation 使用資料篩選來實現資料欄層級安全性、資料列層級安全性和儲存格層級安全性。如果您的來源資料包含巢狀結構，您可以在巢狀資料欄上定義和套用資料篩選條件。

資料欄層級篩選的備註和限制

指定資料欄篩選有三種方式：

- 使用資料篩選條件
- 使用簡單的資料欄篩選或巢狀資料欄篩選。
- 使用 TAGs

簡單資料欄篩選只會指定要包含或排除的資料欄清單。Lake Formation 主控台、API 和 都 AWS CLI 支援簡單的資料欄篩選。如需範例，請參閱「Grant with Simple Column Filtering」。

下列備註和限制適用於資料欄篩選：

- AWS Glue 5.0 或更新版本僅針對 Apache Hive 和 Apache Iceberg 資料表支援透過 Lake Formation 進行精細存取控制。
- 若要SELECT使用授予選項和資料欄篩選進行授予，您必須使用包含清單，而不是排除清單。如果沒有授予選項，您可以使用包含或排除清單。
- 若要SELECT在具有資料欄篩選的資料表上授予 ，您必須已在具有授予選項且沒有任何資料列限制的資料表SELECT上授予 。您必須能夠存取所有資料列。

- 如果您SELECT使用授予選項和資料欄篩選授予您帳戶中的委託人，該委託人必須在授予另一個委託人時為相同的資料欄或授予的資料欄子集指定資料欄篩選。如果您SELECT使用授予選項和資料欄篩選授予外部帳戶，則外部帳戶中的資料湖管理員可以將所有資料欄SELECT上的 授予其帳戶中的另一個主體。不過，即使所有資料欄SELECT都使用 ，該委託人仍只能查看授予外部帳戶的資料欄。
- 您無法在分割區索引鍵上套用資料欄篩選。
- 對資料表中的資料欄子集具有 SELECT許可的主體，無法獲得該資料表上的 ALTER、DELETE、DROP或 INSERT許可。對於資料表上具有 ALTER、DELETE、 DROP或 INSERT許可的委託人，如果您授予具有資料欄篩選的SELECT許可，則不會有任何影響。

下列備註和限制適用於巢狀資料欄篩選：

- 您可以在資料篩選條件中包含或排除五層巢狀欄位。

  Example

  Col1.Col1_1.Col1_1_1.Col1_1_1_1_1_1.Col1_1_1_1_1_1
- 您無法對分割區資料欄內的巢狀欄位套用資料欄篩選。
- 如果您的資料表結構描述包含最上層資料欄名稱 ("customer"."address")，其資料篩選條件內的巢狀欄位表示法模式相同（具有最上層資料欄名稱customer和巢狀欄位名稱的巢狀資料欄address在資料篩選條件"customer"."address"中指定為），則您無法明確指定對最上層資料欄或巢狀欄位的存取，因為兩者在包含/排除清單中都使用相同的模式表示。這是不明確的，而且如果您指定最上層資料欄或巢狀欄位，Lake Formation 無法解析。
- 如果頂層資料欄或巢狀欄位在名稱中包含雙引號，當您在資料儲存格篩選條件的包含和排除清單中指定對巢狀欄位的存取時，您必須包含第二個雙引號。

  Example

  具有雙引號的範例巢狀資料欄名稱 – `a.b.double"quote`

  Example

  資料篩選條件中的巢狀資料欄表示範例 – `"a"."b"."double""quote"`

儲存格層級篩選限制

請記住下列有關資料列層級和儲存格層級篩選的注意事項和限制。

- 巢狀資料欄、檢視和資源連結不支援儲存格層級安全性。

- 巢狀資料欄也支援頂層資料欄上支援的所有表達式。不過，定義巢狀資料列層級表達式時，不應參考分割區資料欄下的巢狀欄位。
- 使用 Athena 引擎第 3 版或 Amazon Redshift Spectrum 時，所有區域都可使用儲存格層級安全性。對於其他 服務，儲存格層級安全性僅適用於 所述的區域支援地區。
- 不支援 SELECT INTO 陳述式。
- 資料列篩選條件表達式不支援 array、 和 map資料類型。支援 struct資料類型。
- 資料表上可定義的資料篩選條件數目沒有限制，但資料表上單一主體的資料篩選條件SELECT許可有 100 個限制。
- 資料表上可包含在授予中的資料篩選條件數目上限為 100。
- 若要使用資料列篩選條件表達式套用資料篩選條件，您必須在所有資料表資料欄上使用 SELECT 授予選項。當授予外部帳戶時，此限制不適用於外部帳戶中的管理員。
- 如果委託人是群組的成員，並且委託人和群組都獲得一部分資料列的許可，則委託人的有效資料列許可是委託人許可和群組許可的聯集。
- 下列資料欄名稱在資料表中受到資料列層級和儲存格層級篩選的限制：
  - ctid
  - oid
  - xmin
  - cmin
  - xmax
  - cmax
  - Tableoid
  - insertxid
  - deletexid
  - importoid
  - redcatuniqueid
- 如果您將全列篩選條件表達式與述詞的其他篩選條件表達式同時套用至資料表，則全列運算式將優先於所有其他篩選條件表達式。
- 當將資料列子集的許可授予外部 AWS 帳戶，而外部帳戶的資料湖管理員將這些許可授予該帳戶中的委託人時，委託人的有效篩選述詞是帳戶述詞和直接授予委託人的任何述詞的交集。

  例如，如果帳戶具有具有述詞的資料列許可，dept='hr'且委託人單獨獲得 的許可country='us'，則委託人只能存取具有 dept='hr'和 的資料列country='us'。

如需儲存格層級篩選的詳細資訊，請參閱 Lake Formation 中的資料篩選和儲存格層級安全性。

如需使用 Amazon Redshift Spectrum 搭配資料列層級安全政策查詢資料表時的考量和限制，請參閱《Amazon Redshift 資料庫開發人員指南》中的使用 RLS 政策的考量和限制。

混合存取模式的考量和限制

混合存取模式可讓您靈活地選擇性地為 中的資料庫和資料表啟用 Lake Formation 許可 AWS Glue Data Catalog。
使用混合存取模式時，您現在有一個增量路徑，可讓您為一組特定使用者設定 Lake Formation 許可，而不會中斷其他現有使用者或工作負載的許可政策。

下列考量和限制適用於混合存取模式。

限制

- 更新 Amazon S3 位置註冊 – 您無法編輯使用服務連結角色向 Lake Formation 註冊的位置參數。
- 使用 LF-Tags 時選擇加入選項 – 當您可以使用 LF-Tags 授予 Lake Formation 許可時，您可以選擇加入委託人，透過選擇已連接 LF-Tags 的資料庫和資料表，強制執行 Lake Formation 許可作為連續步驟。
- 混合存取模式存取 – 在 Lake Formation 中存取混合存取模式僅限於具有資料湖管理員或唯讀管理員許可的使用者。
- 選擇加入主體 – 目前，只有資料湖管理員角色可以選擇加入主體至資源。
- 選擇加入資料庫中的所有資料表 – 在跨帳戶授予中，當您授予許可，並選擇加入資料庫中的所有資料表時，您也需要選擇加入資料庫，才能讓許可運作。

考量事項

- 將向 Lake Formation 註冊的 Amazon S3 位置更新為混合存取模式 – 我們不建議將已向 Lake Formation 註冊的 Amazon S3 資料位置轉換為混合存取模式，即使可以完成此操作。
- 在混合存取模式中註冊資料位置時的 API 行為
  - CreateTable – 無論混合存取模式旗標為何，該位置都被視為已向 Lake Formation 註冊，並選擇加入狀態。因此，使用者需要資料位置許可才能建立資料表。
  - CreatePartition/BatchCreatePartitions/UpdatePartitions （當分割區位置更新為指向以混合模式註冊的位置時） – 無論混合存取模式旗標和選擇加入狀態，Amazon S3 位置都被視為已向 Lake Formation 註冊。因此，使用者需要資料位置許可才能建立或更新資料庫。

- CreateDatabase/UpdateDatabase（當資料庫位置更新以指向在混合存取模式中註冊的位置時）– 無論混合存取模式旗標和選擇加入狀態，該位置都被視為已向 Lake Formation 註冊。因此，使用者需要資料位置許可才能建立或更新資料庫。
- UpdateTable（更新資料表位置以指向在混合存取模式中註冊的位置時）– 無論混合存取模式旗標和選擇加入狀態，該位置都被視為已向 Lake Formation 註冊。因此，使用者需要資料位置許可才能更新資料表。如果資料表位置未更新或指向未向 Lake Formation 註冊的位置，則使用者不需要資料位置許可來更新資料表。

將 Amazon Redshift 資料倉儲資料帶入 的限制 AWS Glue Data Catalog

您可以使用 為 Amazon Redshift 資料倉儲中的分析資料編製目錄和管理存取權 AWS Glue Data Catalog。有下列限制：

- 不支援跨不同 對聯合目錄授予 Lake Formation AWS 帳戶許可。
- 您必須擁有跨帳戶版本設定第 4 AWS 帳戶版，才能跨 共用聯合目錄中的資料庫或資料表。
- Data Catalog 僅支援建立最上層目錄。
- 您只能更新 Redshift 受管儲存 (RMS) 中目錄的描述。
- 對於具有 Redshift 作為儲存位置的目錄、資料庫、資料表，不支援授予許可的 LF 標籤型存取控制 (LF-TBAC) 方法。
- 不支援將聯合目錄以及聯合目錄中的資料庫和資料表的許可設定為`IAMAllowedPrincipals`群組。

- 不支援 Athena、Amazon EMR Spark 等引擎目錄上的資料定義語言 (DDL) 操作，包括設定目錄組態。
- 不支援使用 Athena 在 RMS 資料表上執行 DDL 操作。
- 不支援建立具體化視觀表，無論是透過 Athena、Apache Spark AWS Glue Data Catalog、 或 Amazon Redshift 取用者。
- Athena 不支援多目錄體驗。它一次只能連接到單一的特定目錄。Athena 無法同時存取或查詢多個目錄。
- 不支援透過 Athena 和 Amazon Redshift 在 Iceberg 資料表上標記和分支操作。
- 不支援 RMS 資料表上的時間歷程。
- 不支援具有資料湖資料表的多層級目錄。存放在 Amazon S3 中與資料湖資料表搭配使用的所有資料都必須位於預設值中 AWS Glue Data Catalog，且無法組織成多層目錄。

- 在 Amazon Redshift 中，資料共用不會新增至已註冊的命名空間。叢集和命名空間是同義詞，一旦將叢集發佈到 AWS Glue Data Catalog，您就無法新增資料。
- Amazon EMR on EC2 不支援跨 RMS 資料表和 Amazon S3 資料表聯結。只有 EMR Serverless 支援此功能。
- 不支援外部結構描述和資料表。
- RMS 資料表只能從 Iceberg REST Catalog 中的 AWS Glue 延伸端點存取。
- Hive 資料表無法從連線至 AWS Glue Iceberg REST Catalog 的第三方引擎存取。
- 支援透過 Spark 的 RMS 資料表上 read_committed 隔離層級。
- Redshift 資料庫名稱在 中被視為不區分大小寫 AWS Glue Data Catalog，限制為 128 個字元，可以是帶有破折號 (-) 和底線 (_) 的英數字元。
- 目錄名稱不區分大小寫，限制為 50 個字元，可以是帶有破折號 (-) 和底線 (_) 的英數字元。
- Amazon Redshift 不支援使用 Lake Formation SQL 樣式 GRANT 和 REVOKE 命令來管理發佈至 之資料表的存取許可 AWS Glue Data Catalog。
- 不會強制執行連接到生產者 （來源） Amazon Redshift 叢集的資料列層級安全和動態資料遮罩政策。相反地，Lake Formation 中定義的存取許可將在共用資料上強制執行。
- 不支援在資料表連結上執行資料定義語言 (DDL) 和資料處理語言 (DML) 操作。
- 如果未正確逸出預留關鍵字，則會導致失敗或錯誤。
- 不支援在多目錄案例中加密資料。

S3 資料表目錄整合限制

您可以將 Amazon S3 資料表儲存貯體和資料表與 AWS Glue Data Catalog （資料目錄） 整合，並從 Lake Formation 主控台或使用服務 APIs 將目錄註冊為 Lake Formation 資料位置。

下列限制適用於將 S3 資料表目錄與 Data Catalog 和 Lake Formation 整合：

- 目錄、資料庫和資料表的`S3 Tables`儲存位置不支援以 LF 標籤為基礎的存取控制 (LF-TBAC) 方法來授予許可。
- AWS Glue 和 Lake Formation 不支援混合大小寫資料欄名稱，並將所有資料欄名稱轉換為小寫。轉換為小寫時，您必須確認資料表資料欄名稱是唯一的。使用 `customer_id`而非 `customerId`。只有在預覽版本期間才支援使用混合大小寫資料欄名稱。
- CreateCatalog API 無法在 Amazon S3 中建立資料表儲存貯體。
- SearchTables API 無法搜尋 S3 資料表。

Hive 中繼資料存放區資料共用的考量和限制

使用 AWS Glue Data Catalog 中繼資料聯合 (Data Catalog 聯合），您可以將 Data Catalog 連線到外部中繼存放區，以存放 Amazon S3 資料的中繼資料，並使用 安全地管理資料存取許可 AWS Lake Formation。

下列考量和限制適用於從 Hive 資料庫建立的聯合資料庫：

考量事項

- AWS SAM 應用程式支援 – 您要負責部署的應用程式資源可用性 AWS SAM (Amazon API Gateway 和 Lambda 函數）。當使用者執行查詢時，請確定 AWS Glue Data Catalog 與 Hive 中繼存放區之間的連線正常運作。
- Hive 中繼存放區版本需求 – 您只能使用 Apache Hive 第 3 版及更高版本建立聯合資料庫。
- 映射的資料庫需求 – 每個 Hive 資料庫都必須映射至 Lake Formation 中的新資料庫。
- 資料庫層級聯合支援 – 您只能在資料庫層級連線至 Hive 中繼存放區。
- 聯合資料庫的許可 – 即使刪除來源資料表或資料庫，在聯合資料庫下套用至聯合資料庫或資料表的許可仍會保留。重新建立來源資料庫或資料表時，您不需要重新授予許可。在來源刪除具有 Lake Formation 許可的聯合資料表時，Lake Formation 許可仍然可見，您可以視需要撤銷它們。

  如果使用者刪除聯合資料庫，則會失去其所有對應的許可。以相同名稱重新建立相同的資料庫， 不會復原 Lake Formation 許可。使用者必須再次設定新的許可。
- 聯合資料庫上的 IAMAllowedPrincipal 群組許可 – 根據 `DataLakeSettings`，Lake Formation 可能會將所有資料庫和資料表的許可設定為名為 的虛擬群組`IAMAllowedPrincipal`。`IAMAllowedPrincipal` 是指可透過 IAM 主體政策和資源政策存取 Data Catalog AWS Glue 資源的所有 IAM 主體。如果這些許可存在於資料庫或資料表上，則會授予所有主體存取資料庫或資料表的權限。

  不過，Lake Formation 不允許對聯合資料庫下的資料表進行`IAMAllowedPrincipal`許可。當您建立聯合資料庫時，請確定您將 `CreateTableDefaultPermissions` 參數做為空清單傳遞。

  如需詳細資訊，請參閱變更資料湖的預設設定。
- 在查詢中聯結資料表 – 您可以將 Hive 中繼存放區資料表與 Data Catalog 原生資料表聯結，以執行查詢。

限制

- AWS Glue Data Catalog 與 Hive 中繼存放區之間同步中繼資料的限制 – 建立 Hive 中繼存放區連線後，您需要建立聯合資料庫，以將 Hive 中繼存放區中的中繼資料與 同步 AWS Glue Data Catalog。當使用者執行查詢時，聯合資料庫下的資料表會在執行時間同步。
- 在聯合資料庫下建立新資料表的限制 – 您將無法在聯合資料庫下建立新資料表。
- 資料許可限制 – 不支援 Hive 中繼存放區資料表檢視的許可。

Amazon Redshift 資料共用限制

AWS Lake Formation 可讓您從 Amazon Redshift 安全地管理資料共用中的資料。Amazon Redshift 是 AWS 雲端中全受管的 PB 級資料倉儲服務。使用資料共用功能，Amazon Redshift 可協助您跨 共用資料 AWS 帳戶。如需 Amazon Redshift 資料共用的詳細資訊，請參閱 Amazon Redshift 中的資料共用概觀。

下列備註和限制適用於從 Amazon Redshift 資料共用建立的聯合資料庫：

- 映射的資料庫需求 – 每個 Amazon Redshift 資料共用都必須映射至 Lake Formation 中的新資料庫。在 Data Catalog 資料庫中扁平化資料共用物件表示法時，這是維持唯一資料表名稱的必要項目。
- 在聯合資料庫下建立新資料表的限制 – 您將無法在聯合資料庫下建立新資料表。
- 聯合資料庫上的許可 – 即使刪除來源資料表或資料庫，在聯合資料庫下套用至聯合資料庫或資料表的許可仍會保留。重新建立來源資料庫或資料表時，您不需要重新授予許可。在來源刪除具有 Lake Formation 許可的聯合資料表時，Lake Formation 許可仍然可見，您可以視需要撤銷它們。

  如果使用者刪除聯合資料庫，則會失去其所有對應的許可。以相同名稱重新建立相同的資料庫， 不會復原 Lake Formation 許可。使用者必須再次設定新的許可。
- 聯合資料庫上的 IAMAllowedPrincipal 群組許可 – 根據 `DataLakeSettings`，Lake Formation 可能會將所有資料庫和資料表的許可設定為名為 的虛擬群組`IAMAllowedPrincipal`。`IAMAllowedPrincipal` 是指可透過 IAM 主體政策和資源政策存取 Data Catalog AWS Glue 資源的所有 IAM 主體。如果這些許可存在於資料庫或資料表上，則會授予所有主體存取資料庫或資料表的權限。

  不過，Lake Formation 不允許對聯合資料庫下的資料表進行`IAMAllowedPrincipal`許可。當您建立聯合資料庫時，請確定您將 `CreateTableDefaultPermissions` 參數做為空清單傳遞。

  如需詳細資訊，請參閱變更資料湖的預設設定。

- 資料篩選 – 在 Lake Formation 中，您可以在具有資料欄層級和資料列層級篩選的聯合資料庫下授予資料表的許可。不過，您無法合併資料欄層級和資料列層級篩選，以限制聯合資料庫下資料表的儲存格層級精細程度存取。
- 區分大小寫識別符 – Lake Formation 管理的 Amazon Redshift 資料共用物件僅支援小寫的資料表名稱和資料欄名稱。如果資料庫、資料表和資料欄將使用 Lake Formation 共用和管理，請勿開啟 Amazon Redshift 資料共用中的區分大小寫識別符。
- 查詢支援 – 您可以使用 Amazon Redshift 查詢 Lake Formation 管理的 Amazon Redshift 資料共用。Athena 不支援查詢 Lake Formation 管理的 Amazon Redshift 資料共用。

如需在 Amazon Redshift 中使用資料共用時限制的詳細資訊，請參閱《Amazon Redshift 資料庫開發人員指南》中的資料共用限制。

IAM Identity Center 整合限制

使用 AWS IAM Identity Center，您可以連線到身分提供者 IdPs)，並跨 AWS 分析服務集中管理使用者和群組的存取權。您可以在 IAM Identity Center 中將 AWS Lake Formation 設定為已啟用的應用程式，而資料湖管理員可以將精細許可授予 AWS Glue Data Catalog 資源上的授權使用者和群組。

下列限制適用於 Lake Formation 與 IAM Identity Center 的整合：

- 您無法在 Lake Formation 中將 IAM Identity Center 使用者和群組指派為資料湖管理員或唯讀管理員。

  如果您使用 AWS Glue 可代表您擔任的 IAM 角色來加密和解密 Data Catalog，IAM Identity Center 使用者和群組可以查詢加密的 Data Catalog 資源。 AWS 受管金鑰不支援受信任的身分傳播。
- IAM Identity Center 使用者和群組只能叫用 IAM Identity Center 所提供`AWSIAMIdentityCenterAllowListForIdentityContext`政策中列出的 API 操作。
- Lake Formation 允許來自外部帳戶的 IAM 角色代表 IAM Identity Center 使用者和群組擔任電信業者角色，以存取 Data Catalog 資源，但只能在擁有帳戶內的 Data Catalog 資源上授予許可。如果您嘗試將許可授予外部帳戶中 Data Catalog 資源上的 IAM Identity Centerusers 和群組，Lake Formation 會擲出下列錯誤 -「委託人不支援跨帳戶授予。」

Lake Formation 標籤型存取控制最佳實務和考量事項

您可以建立、維護和指派 LF 標籤，以控制對 Data Catalog 資料庫、資料表和資料欄的存取。

使用 Lake Formation 標籤型存取控制時，請考慮下列最佳實務：

- 所有 LF 標籤都必須預先定義，才能指派給 Data Catalog 資源或授予委託人。

  資料湖管理員可以透過建立具有所需 IAM 許可的 LF-Tag 建立者來委派標籤管理任務。資料工程師和分析師決定 LF 標籤的特性和關係。然後，LF 標籤建立者會在 Lake Formation 中建立和維護 LF 標籤。

- 您可以將多個 LF 標籤指派給 Data Catalog 資源。只能將特定金鑰的一個值指派給特定資源。

  例如，您可以將 `module=Orders`、`division=Consumer`、 `region=West`等指派給資料庫、資料表或資料欄。您無法指派 `module=Orders,Customers`。

- 您無法在建立資源時將 LF 標籤指派給資源。您只能將 LF 標籤新增至現有資源。

- 您可以授予 LF-Tag 表達式給委託人，而不只是單一 LF-Tag。

  LF-Tag 表達式看起來類似以下內容 （虛擬程式碼）。

  ```
  module=sales AND division=(consumer OR commercial)
  ```

  授予此 LF-Tag 表達式的主體只能存取已指派`module=sales`和 `division=consumer`或 的資料目錄資源 （資料庫、資料表和資料欄）`division=commercial`。如果您希望委託人能夠存取具有 `module=sales`或 的資源`division=commercial`，請不要將兩者都包含在相同的授予中。進行兩次授予，一次用於 `module=sales` ，另一次用於 `division=commercial`。

  最簡單的 LF-Tag 表達式僅包含一個 LF-Tag，例如 `module=sales`。

- 在具有多個值的 LF-Tag 上授予許可的主體可以使用其中一個值存取 Data Catalog 資源。例如，如果授予使用者 LF 標籤，且金鑰 `=module` 且值 `=orders,customers`，則使用者可以存取指派給 `module=orders`或 的資源`module=customers`。

- 您需要具有`Grant with LF-Tag expressions`許可，才能使用 LF-TBAC 方法授予 Data Catalog 資源的資料許可。資料湖管理員和 LF 標籤建立者會隱含地接收此許可。具有 `Grant with LFTag expressions`許可的主體可以使用下列方式授予資源的資料許可：

  - 具名資源方法
  - LF-TBAC 方法，但僅使用相同的 LF-Tag 表達式

    例如，假設資料湖管理員進行下列授予 （虛擬程式碼）。

    ```
    GRANT (SELECT ON TABLES) ON TAGS module=customers, region=west,south TO user1 WITH
     GRANT OPTION
    ```

在此情況下， user1可以使用 LF-TBAC 方法將資料表SELECT上的 授予其他主體，但只能使用完整的 LF-Tag 表達式 module=customers, region=west,south。

- 如果委託人獲得同時具有 LF-TBAC 方法和具名資源方法的資源許可，則委託人在資源上擁有的許可是兩種方法所授予許可的聯集。
- Lake Formation 支援跨帳戶授予 DESCRIBE 和 ASSOCIATE LF 標籤，以及使用 LF-TBAC 方法跨帳戶授予 Data Catalog 資源的許可。在這兩種情況下，委託人都是 AWS 帳戶 ID。

> **Note**
>
> Lake Formation 支援使用 LF-TBAC 方法跨帳戶授予組織和組織單位。若要使用此功能，您需要將跨帳戶版本設定更新為第 3 版。

如需詳細資訊，請參閱Lake Formation 中的跨帳戶資料共用。

- 在一個帳戶中建立的資料目錄資源只能使用相同帳戶中建立的 LF 標籤進行標記。在一個帳戶中建立的 LF 標籤無法與另一個帳戶的共用資源建立關聯。
- 使用 Lake Formation 標籤型存取控制 (LF-TBAC) 授予 Data Catalog 資源的跨帳戶存取權，需要為 AWS 您的帳戶新增 Data Catalog 資源政策。如需詳細資訊，請參閱先決條件。
- LF 標籤索引鍵和 LF 標籤值的長度不能超過 50 個字元。
- 可指派給 Data Catalog 資源的 LF 標籤數目上限為 50。
- 下列限制為軟性限制：
  - 可建立的 LF 標籤數目上限為 1000。
  - LF 標籤可定義的值數目上限為 1000。
- 標籤索引鍵和值會在儲存時轉換為所有小寫。
- 只有一個 LF-Tag 的值可以指派給特定資源。
- 如果以單一授權將多個 LF 標籤授予委託人，委託人只能存取具有所有 LF 標籤的 Data Catalog 資源。
- 如果 LF-Tag 表達式評估僅導致存取一部分的資料表資料欄，但符合時授予的 Lake Formation 許可是需要完整資料欄存取的許可之一，即 Alter、Insert、 Drop或 Delete，則不會授予任何這些許可。反Describe之，只會授予 。如果授予的許可是 All(Super)，則只會授予 Describe Select和 。

- 萬用字元不會與 LF 標籤搭配使用。若要將 LF-Tag 指派給資料表的所有資料欄，您可以將 LF-Tag 指派給資料表，而資料表中的所有資料欄都會繼承 LF-Tag。若要將 LF-Tag 指派給資料庫中的所有資料表，您可以將 LF-Tag 指派給資料庫，而資料庫中的所有資料表都會繼承該 LF-Tag。
- 您可以在帳戶中建立最多 1000 個 LF-Tag 表達式。
- 您可以使用最多 50 個 LF-Tag 表達式，將許可授予 Data Catalog 資源上的委託人。
- 在內嵌 LF-Tag 表達式上授予或撤銷許可時，LF-Tag 表達式的大小不得超過 900 個位元組。若要授予較大 LF-Tag 表達式的許可，請使用儲存的 LF-Tag 表達式。如需詳細資訊，請參閱建立 LF-Tag 表達式。

屬性型存取控制考量事項、限制和支援的區域

下列考量和限制適用於屬性型存取控制 (ABAC)。

- ABAC 不支援使用 LF-Tag 政策授予存取權。
- ABAC 不提供可授予的許可。
- ABAC 不支援將許可授予 IAM Identity Center 使用者。
- 在 Lake Formation 的資料表上使用 ABAC 授權時，Lake Formation 不會將`DESCRIBE`許可授予父資料庫或目錄。這與非 ABAC 案例不同，其中 Lake Formation 提供父資源的隱含`DESCRIBE`許可。
- 具有`AmazonDataZoneProject`標籤索引鍵的所有主體一律視為已針對所有資料目錄資源選擇加入 Lake Formation。
- ABAC 僅支援字串屬性。

Lake Formation 故障診斷

如果您在使用 AWS Lake Formation 時遇到問題，請參閱本節中的主題。

主題

- 一般性問題的故障診斷
- 對跨帳戶存取進行故障診斷
- 疑難排解藍圖和工作流程
- 的已知問題 AWS Lake Formation
- 已更新錯誤訊息

一般性問題的故障診斷

使用此處的資訊來協助您診斷和修正各種 Lake Formation 問題。

錯誤：<Amazon S3 location> 上的 Lake Formation 許可不足

嘗試在資源指向的 Amazon S3 位置上建立或修改 Data Catalog 資源，而沒有資料位置許可。

如果 Data Catalog 資料庫或資料表指向 Amazon S3 位置，當您授予 Lake Formation 許可`CREATE_TABLE`或 時`ALTER`，也必須授予該位置的`DATA_LOCATION_ACCESS`許可。如果您要將這些許可授予外部帳戶或組織，則必須包含授予選項。

將這些許可授予外部帳戶後，該帳戶中的資料湖管理員必須將許可授予帳戶中的主體 （使用者或角色）。授予從另一個帳戶收到的`DATA_LOCATION_ACCESS`許可時，您必須指定擁有者帳戶的目錄 ID (AWS 帳戶 ID)。擁有者帳戶是註冊位置的帳戶。

如需詳細資訊，請參閱 基礎資料存取控制 和 授予資料位置許可。

錯誤：「Glue API 的加密金鑰許可不足」

已嘗試授予 Lake Formation 許可，而沒有 AWS Identity and Access Management (IAM) 許可給加密 Data Catalog 的 AWS KMS 加密金鑰。

使用資訊清單的我 Amazon Athena 或 Amazon Redshift 查詢失敗

Lake Formation 不支援使用資訊清單的查詢。

錯誤：「Lake Formation (Lake Formation) 許可不足：目錄上的必要建立標籤」

使用者/角色必須是資料湖管理員。

刪除無效的資料湖管理員時發生錯誤

您應該同時刪除所有無效的資料湖管理員（已刪除定義為資料湖管理員的 IAM 角色）。如果您嘗試分別刪除無效的資料湖管理員，Lake Formation 會擲回無效的主體錯誤。

對跨帳戶存取進行故障診斷

使用此處的資訊來協助您診斷和修正跨帳戶存取問題。

主題

- 我授予跨帳戶 Lake Formation 許可，但收件人看不到資源
- 收件人帳戶中的主體可以查看 Data Catalog 資源，但無法存取基礎資料
- 錯誤：接受 AWS RAM 資源共享邀請時，「關聯失敗，因為發起人未授權」
- 錯誤：「未授權授予資源的許可」
- 錯誤：「拒絕擷取 AWS 組織資訊的存取」
- 錯誤：「找不到組織 <organization-ID>」
- 錯誤：「Lake Formation 許可不足：組合無效」
- 對外部帳戶授予/撤銷請求的 ConcurrentModificationException
- 使用 Amazon EMR 存取跨帳戶共用的資料時發生錯誤

我授予跨帳戶 Lake Formation 許可，但收件人看不到資源

- 收件人帳戶中的使用者是否為資料湖管理員？ 只有資料湖管理員可以在共用時查看資源。
- 您是否使用具名資源方法與組織外部的帳戶共用？ 若是如此，收件人帳戶的資料湖管理員必須在 AWS Resource Access Manager () 中接受資源共享邀請AWS RAM。

  如需詳細資訊，請參閱the section called “接受 AWS RAM 資源共享邀請”。
- 您是否在 中使用帳戶層級（資料目錄）資源政策AWS Glue？ 如果是，則如果您使用具名資源方法，則必須在 政策中包含一個特殊陳述式，授權 代表您 AWS RAM 共用政策。

如需詳細資訊，請參閱the section called “使用 AWS Glue和 Lake Formation 管理跨帳戶許可”。

- 您是否有授予跨帳戶存取權所需的 AWS Identity and Access Management (IAM) 許可？

  如需詳細資訊，請參閱the section called “先決條件”。

- 您授予許可的資源不得將任何 Lake Formation 許可授予`IAMAllowedPrincipals`群組。
- 帳戶層級政策中的資源是否有`deny`陳述式？

收件人帳戶中的主體可以查看 Data Catalog 資源，但無法存取基礎資料

收件人帳戶中的委託人必須具有必要的 AWS Identity and Access Management (IAM) 許可。如需詳細資訊，請參閱 存取共用資料表的基礎資料。

錯誤：接受 AWS RAM 資源共享邀請時，「關聯失敗，因為發起人未授權」

將資源的存取權授予不同帳戶後，當接收帳戶嘗試接受資源共享邀請時，動作會失敗。

```
$ 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
        }
    ]
}
```

當接收帳戶接受資源共享邀請AWS Glue時， `glue:PutResourcePolicy` 會叫用 ，因此發生錯誤。若要解決此問題，請允許生產者/授權者帳戶使用的擔任角色執行 `glue:PutResourcePolicy`動作。

錯誤：「未授權授予資源的許可」

已嘗試對另一個帳戶擁有的資料庫或資料表授予跨帳戶許可。當資料庫或資料表與您的帳戶共用時，身為資料湖管理員，您只能將資料庫或資料表的許可授予帳戶中的使用者。

錯誤：「拒絕擷取 AWS 組織資訊的存取」

您的帳戶是 AWS Organizations 管理帳戶，您沒有擷取組織資訊所需的許可，例如帳戶中的組織單位。

如需詳細資訊，請參閱Required permissions for cross-account grants。

錯誤：「找不到組織 <organization-ID>」

嘗試與組織共用資源，但未啟用與組織共用。啟用與組織共用資源。

如需詳細資訊，請參閱AWS RAM 《使用者指南》中的啟用與 AWS Organizations 共用。

錯誤：「Lake Formation 許可不足：組合無效」

當 Lake Formation 許可授予資源的 `IAMAllowedPrincipals` 群組時，使用者共用 Data Catalog 資源。使用者必須先從 撤銷所有 Lake Formation 許可，`IAMAllowedPrincipals`才能共用資源。

對外部帳戶授予/撤銷請求的 ConcurrentModificationException

當使用者對 LF-Tag 政策上的委託人提出多個並行授予和/或撤銷許可請求時，Lake Formation 會擲回 ConcurrentModificationException。使用者需要擷取例外狀況，並重試失敗的授予/撤銷請求。使用 `GrantPermissions`/`RevokePermissions` API 操作的批次版本 - BatchGrantPermissions 和 BatchRevokePermissions 可透過減少並行授予/撤銷請求的數量，在一定程度上緩解此問題。

使用 Amazon EMR 存取跨帳戶共用的資料時發生錯誤

當您使用 Amazon EMR 存取從其他帳戶與您共用的資料時，部分 Spark 程式庫將嘗試呼叫 `Glue:GetUserDefinedFunctions` API 操作。由於 AWS RAM 受管許可的版本 1 和 2 不支援此動作，因此您會收到下列錯誤訊息：

```
"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"
```

若要解決此錯誤，建立資源共用的資料湖管理員必須更新連接至資源共用的 AWS RAM 受管許可。AWS RAM 受管許可第 3 版允許主體執行 `glue:GetUserDefinedFunctions` 動作。

如果您建立新的資源共享，Lake Formation 預設會套用最新版本的 AWS RAM 受管許可，而且您不需要採取任何動作。若要啟用現有資源共享的跨帳戶資料存取，您需要將 AWS RAM 受管許可更新為第 3 版。

您可以在 中檢視指派給與您共用資源的 AWS RAM 許可 AWS RAM。第 3 版中包含下列許可：

```
Databases
  AWSRAMPermissionGlueDatabaseReadWriteForCatalog
  AWSRAMPermissionGlueDatabaseReadWrite

Tables
  AWSRAMPermissionGlueTableReadWriteForCatalog
  AWSRAMPermissionGlueTableReadWriteForDatabase

AllTables
  AWSRAMPermissionGlueAllTablesReadWriteForCatalog
  AWSRAMPermissionGlueAllTablesReadWriteForDatabase
```

更新現有資源共享的 AWS RAM 受管許可版本

您（資料湖管理員）可以依照 AWS RAM 使用者指南中的指示，將AWS RAM 受管許可更新為較新版本，也可以撤銷資源類型的所有現有許可，並予以重新授予。如果您撤銷許可， 會 AWS RAM 刪除與 AWS RAM 資源類型相關聯的資源共用。當您授予許可時， AWS RAM 會建立新的資源共用，以連接最新版本的 AWS RAM 受管許可。

疑難排解藍圖和工作流程

使用此處的資訊來協助您診斷和修正藍圖和工作流程問題。

主題

- 我的藍圖失敗，「使用者：<user-ARN> 未獲授權執行：iam：PassRole on resource： <role-ARN>」
- 我的工作流程失敗，「使用者：<user-ARN> 未獲授權執行：iam：PassRole on resource： <role-ARN>」
- 我工作流程中的爬蟲程式失敗，出現「資源不存在或請求者無權存取請求的許可」

- 工作流程中的爬蟲程式失敗，並顯示「呼叫 CreateTable 操作時發生錯誤 (AccessDeniedException)...」

我的藍圖失敗，「使用者：<user-ARN> 未獲授權執行：iam：PassRole on resource： <role-ARN>」

嘗試建立藍圖的使用者沒有足夠的許可來傳遞所選角色。

更新使用者的 IAM 政策以傳遞角色，或要求使用者選擇具有必要通行許可的不同角色。

如需詳細資訊，請參閱the section called “Lake Formation 角色和 IAM 許可參考”。

我的工作流程失敗，「使用者：<user-ARN> 未獲授權執行：iam：PassRole on resource： <role-ARN>」

您為工作流程指定的角色沒有允許角色自行傳遞的內嵌政策。

如需詳細資訊，請參閱the section called “（選用）建立工作流程的 IAM 角色”。

我工作流程中的爬蟲程式失敗，出現「資源不存在或請求者無權存取請求的許可」

其中一個可能的原因，是傳遞的角色沒有足夠的許可，無法在目標資料庫中建立資料表。授予該角色資料庫的CREATE_TABLE許可。

工作流程中的爬蟲程式失敗，並顯示「呼叫 CreateTable 操作時發生錯誤 (AccessDeniedException)...」

其中一個可能的原因，是工作流程角色在目標儲存位置上沒有資料位置許可。將資料位置許可授予角色。

如需詳細資訊，請參閱the section called “DATA_LOCATION_ACCESS”。

的已知問題 AWS Lake Formation

檢閱 的這些已知問題 AWS Lake Formation。

主題

- 篩選資料表中繼資料的限制
- 重新命名排除資料欄時的問題
- 刪除 CSV 資料表中資料欄的問題
- 資料表分割區必須新增至通用路徑下
- 在工作流程建立期間建立資料庫的問題
- 刪除並重新建立使用者的問題
- Data Catalog API 操作不會更新 IsRegisteredWithLakeFormation 參數的值
- Lake Formation 操作不支援 AWS Glue 結構描述登錄檔

篩選資料表中繼資料的限制

AWS Lake Formation 資料欄層級許可可用來限制對資料表中特定資料欄的存取。當使用者使用 主控台或 等 API 擷取資料表的中繼資料時`glue:GetTable`，資料表物件中的資料欄清單只會包含他們有權存取的欄位。請務必了解此中繼資料篩選的限制。

雖然 Lake Formation 為整合服務提供資料欄許可的中繼資料，但查詢回應中的資料欄實際篩選是整合服務的責任。支援資料欄層級篩選的 Lake Formation 用戶端，包括 Amazon Athena、Amazon Redshift Spectrum 和 Amazon EMR，根據向 Lake Formation 註冊的資料欄許可來篩選資料。使用者將無法讀取他們不應存取的任何資料。目前，AWS GlueETL 不支援資料欄篩選。

> **Note**
>
> EMR 叢集未完全由 管理 AWS。因此，EMR 管理員有責任正確保護叢集，以避免未經授權存取資料。

某些應用程式或格式可能會在`Parameters`映射中將其他中繼資料存放為資料表屬性，包括資料欄名稱和類型。這些屬性會以未修改的方式傳回，並可由任何具有任何資料欄`SELECT`許可的任何使用者存取。

例如，Avro SerDe 會將資料表結構描述的 JSON 表示法存放在名為 的資料表屬性中`avro.schema.literal`，可供有權存取資料表的所有使用者使用。我們建議您避免將敏感資訊存放在資料表屬性中，並請注意，使用者可以了解 Avro 格式資料表的完整結構描述。此限制專屬於資料表的中繼資料。

AWS Lake Formation 如果發起人沒有資料表中所有資料欄的`SELECT`許可，則在回應`glue:GetTable`或類似請求`spark.sql.sources.schema`時， 會移除以 開頭的任何資料表屬

性。這可防止使用者存取使用 Apache Spark 建立之資料表的其他中繼資料。在 Amazon EMR 上執行時，Apache Spark 應用程式仍然可以讀取這些資料表，但可能不會套用某些最佳化，並且不支援區分大小寫的資料欄名稱。如果使用者可存取資料表中的所有資料欄，Lake Formation 會傳回未修改且具有所有資料表屬性的資料表。

重新命名排除資料欄時的問題

如果您使用資料欄層級許可來排除資料欄，然後重新命名資料欄，則資料欄不會再從查詢中排除，例如 `SELECT *`。

刪除 CSV 資料表中資料欄的問題

如果您使用 CSV 格式建立 Data Catalog 資料表，然後從結構描述中刪除資料欄，則查詢可能會傳回錯誤的資料，而且可能未遵守資料欄層級許可。

解決方法：改為建立新的資料表。

資料表分割區必須新增至通用路徑下

Lake Formation 預期資料表的所有分割區都位於資料表位置欄位中設定的常見路徑下。當您使用爬蟲程式將分割區新增至目錄時，這會順暢運作。但是，如果您手動新增分割區，且這些分割區不在父資料表中設定的位置之下，則資料存取無法運作。

在工作流程建立期間建立資料庫的問題

使用 Lake Formation 主控台從藍圖建立工作流程時，如果目標資料庫不存在，您可以建立目標資料庫。當您這麼做時，登入的使用者會取得所建立資料庫的 `CREATE_TABLE` 許可。不過，工作流程產生的爬蟲程式會在嘗試建立資料表時擔任工作流程的角色。這會失敗，因為角色沒有資料庫的 `CREATE_TABLE`許可。

解決方法：如果您在工作流程設定期間透過主控台建立資料庫，則在執行工作流程之前，您必須為與工作流程相關聯的角色提供您剛建立之資料庫的`CREATE_TABLE`許可。

刪除並重新建立使用者的問題

以下案例導致 傳回錯誤的 Lake Formation 許可`lakeformation:ListPermissions`：

1. 建立使用者並授予 Lake Formation 許可。
2. 刪除使用者。

3. 使用相同名稱重新建立使用者。

ListPermissions 會傳回兩個項目，一個用於舊使用者，另一個用於新使用者。如果您嘗試撤銷授予舊使用者的許可，則會撤銷新使用者的許可。

Data Catalog API 操作不會更新 **IsRegisteredWithLakeFormation** 參數的值

Data Catalog API 操作有已知的限制，例如 GetTables和 SearchTables 不會更新 IsRegisteredWithLakeFormation 參數的值，並傳回預設值，這是 false。建議使用 GetTable API 來檢視 IsRegisteredWithLakeFormation 參數的正確值。

Lake Formation 操作不支援 AWS Glue 結構描述登錄檔

Lake Formation 操作不支援在 SchemaReference中包含StorageDescriptor要在結構描述註冊中使用的 的 AWS Glue 資料表。

已更新錯誤訊息

AWS Lake Formation 已將資源特定例外狀況更新為下列 API 操作的一般EntityNotFound錯誤訊息，以符合安全性和合規目標。

- RevokePermissions
- GrantPermissions
- GetResourceLFTags
- GetTable
- GetDatabase

AWS Lake Formation API

> **Note**
>
> AWS Lake Formation 服務的更新 API 參考現已可用。

內容

- 許可 APIs
  - 作業
  - 資料類型
- 資料湖設定 APIs
  - 作業
  - 資料類型
- IAM Identity Center APIs
  - 作業
  - 資料類型
- 混合存取模式 APIs
  - 作業
  - 資料類型
- 登入資料販賣 APIs
  - 作業
  - 資料類型
- 標記 API
  - 作業
  - 資料類型
- 資料篩選條件 APIs
  - 作業
  - 資料類型
- 常見資料類型
  - ErrorDetail 結構

- 字串模式

許可 APIs

許可 API 區段說明授予和撤銷許可所需的操作和資料類型 AWS Lake Formation。如需所有 API 操作和資料類型，請參閱 Lake Formation API 參考指南。 AWS Lake Formation

作業

- GrantPermissions
- RevokePermissions
- BatchGrantPermissions
- BatchRevokePermissions
- GetEffectivePermissionsForPath
- ListPermissions
- GetDataLakePrincipal

資料類型

- Resource
- DatabaseResource
- TableResource
- TableWithColumnsResource
- DataCellsFilterResourcee
- DataLocationResource
- DataLakePrincipal
- PrincipalPermissions
- PrincipalResourcePermissions
- DetailsMap
- ColumnWildcard
- BatchPermissionsRequestEntry
- BatchPermissionsFailureEntry

資料湖設定 APIs

本節包含用於管理資料湖管理員的資料湖設定 API 操作和資料類型。

作業

- GetDataLakeSettings
- PutDataLakeSettings

資料類型

- DataLakeSettings

IAM Identity Center APIs

本節包含建立和管理 Lake Formation 與 IAM Identity Center 整合的操作。

作業

- CreateLakeFormationIdentityCenterConfiguration
- DeleteLakeFormationIdentityCenterConfiguration
- DescribeLakeFormationIdentityCenterConfiguration
- UpdateLakeFormationIdentityCenterConfiguration

資料類型

- ExternalFilteringConfiguration

混合存取模式 APIs

混合存取模式 API 區段說明在 中設定混合存取模式所需的操作和資料類型 AWS Lake Formation。如需所有 API 操作和資料類型，請參閱 Lake Formation API 參考指南。 AWS Lake Formation

作業

- CreateLakeFormationOptIn

- DeleteLakeFormationOptIn
- ListLakeFormationOptIns

資料類型

- Resource
- DatabaseResource
- TableResource
- 資源資訊
- LakeFormationOptInsInfo
- DataLocationResource

登入資料販賣 APIs

登入資料販賣 API 區段說明與使用 AWS Lake Formation 服務相關的操作和資料類型，以便提供登入資料，以及註冊和管理資料湖資源。

作業

- RegisterResource
- DeregisterResource
- ListResources
- GetUnfilteredTableMetadata
- GetUnfilteredPartitionsMetadata
- GetTemporaryGluePartitionCredentials
- GetTemporaryGlueTableCredentials
- UpdateResource

資料類型

- FilterCondition
- RowFilter

- ResourceInfo

標記 API

標記 API 區段說明與授權策略相關的操作和資料類型，該策略定義屬性或鍵值對標籤的許可模型。

作業

- GetLFTagExpression
- ListLFTagExpressions
- DeleteLFTagExpression
- UpdateLFTagExpression
- CreateLFTagExpression
- AddLFTagsToResource
- RemoveLFTagsFromResource
- GetResourceLFTags
- ListLFTags
- CreateLFTag
- GetLFTag
- UpdateLFTag
- DeleteLFTag
- SearchTablesByLFTags
- SearchDatabasesByLFTags

資料類型

- LFTagKeyResource
- LFTagPolicyResource
- TaggedTable
- TaggedDatabase
- LFTag
- LFTagPair

- LFTagError
- ColumnLFTag

資料篩選條件 APIs

資料篩選條件 APIs 說明如何管理其中的資料儲存格篩選條件 AWS Lake Formation。

作業

- CreateDataCellsFilter
- DeleteDataCellsFilter
- ListDataCellsFilter
- GetDataCellsFilter
- UpdateDataCellsFilter

資料類型

- DataCellsFilter
- RowFilter

常見資料類型

Common Data Types 說明 AWS Lake Formation中的其他常見資料類型。

ErrorDetail 結構

包含錯誤的詳細資訊。

欄位

- `ErrorCode` – UTF-8 字串，長度不可小於 1 個位元組，也不可以超過 255 個位元組，需符合Single-line string pattern。

  此錯誤相關的程式碼。
- `ErrorMessage` – 描述字串，長度不可超過 2048 個位元組，需符合URI address multi-line string pattern。

描述錯誤的訊息。

字串模式

API 使用以下常規表達式來定義適用於各種字串參數和成員的有效內容：

- 單行字串模式 –「[\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\t]*」
- URI 位址多行字串模式 –「[\u0020-\uD7FF\uE000-\uFFFD\uD800\uDC00-\uDBFF\uDFFF\r\n\t]*」
- 自訂字串模式 #3 – "^\w+\.\w+\.\w+$"
- 自訂字串模式 #4 – "^\w+\.\w+$"
- 自訂字串模式 #5 – "arn:aws:iam::[0-9]*:role/.*"
- 自訂字串模式 #6 – "arn:aws:iam::[0-9]*:user/.*"
- 自訂字串模式 #7 – "arn:aws:iam::[0-9]*:group/.*"
- 自訂字串模式 #8 – "arn:aws:iam::[0-9]*:saml-provider/.*"
- 自訂字串模式 #9 – "^([\p{L}\p{Z}\p{N}_.:\/=+\-@%]*)$"
- 自訂字串模式 #10 – "^([\p{L}\p{Z}\p{N}_.:*\/=+\-@%]*)$"
- 自訂字串模式 #11 – "[\p{L}\p{N}\p{P}]*"

支援地區

本節提供有關 Lake Formation 支援 AWS 區域 和 功能的資訊。

一般可用性

如需 AWS 區域 支援的 AWS Lake Formation，請參閱區域可用的 AWS 服務清單。

如需每個區域和 Lake Formation 服務配額的 Lake Formation 服務端點清單，請參閱AWS Lake Formation 端點和配額。

AWS GovCloud (US)

如需 AWS GovCloud (US) 區域與標準 間差異的概觀 AWS 區域，請參閱 AWS Lake Formation 的不同之處 AWS GovCloud (US)。

交易和儲存最佳化

Lake Formation 的受管資料表、交易支援和儲存最佳化功能提供如下 AWS 區域：

區域名稱	區域參數	端點
美國東部 (維吉尼亞北部)	us-east-1	lakeformation.us-east-1.amazonaws.com lakeformation-fips.us-east-1.amazonaws.com
美國東部 (俄亥俄)	us-east-2	lakeformation.us-east-2.amazonaws.com lakeformation-fips.us-east-2.amazonaws.com
美國西部 (奧勒岡)	us-west-2	lakeformation.us-west-2.amazonaws.com

區域名稱	區域參數	端點
		lakeformation-fips.us-west-2.amazonaws.com
亞太區域 (孟買)	ap-south-1	lakeformation.ap-south-1.amazonaws.com
亞太區域 (首爾)	ap-northeast-2	lakeformation.ap-northeast-2.amazonaws.com
亞太區域 (新加坡)	ap-southeast-1	lakeformation.ap-southeast-1.amazonaws.com
亞太區域 (雪梨)	ap-southeast-2	lakeformation.ap-southeast-2.amazonaws.com
亞太區域 (東京)	ap-northeast-1	lakeformation.ap-northeast-1.amazonaws.com
歐洲 (法蘭克福)	eu-central-1	lakeformation.eu-central-1.amazonaws.com
歐洲 (愛爾蘭)	eu-west-1	lakeformation.eu-west-1.amazonaws.com
歐洲 (倫敦)	eu-west-2	lakeformation.eu-west-2.amazonaws.com
Europe (Stockholm)	eu-north-1	lakeformation.eu-north-1.amazonaws.com
加拿大 (中部)	ca-central-1	lakeformation.ca-central-1.amazonaws.com

區域名稱	區域參數	端點
南美洲 (聖保羅)	sa-east-1	lakeformation.sa-east-1.amazonaws.com

的文件歷史記錄 AWS Lake Formation

下表說明 文件的重要變更 AWS Lake Formation。

變更	描述	日期
更新屬性型存取控制的文件	Lake Formation 可讓您將許可授予在 Data Catalog 資源上具有相符屬性的主體。如需詳細資訊，請參閱屬性型存取控制。	2025 年 4 月 17 日
Amazon S3 Tables 與 AWS Lake Formation 和 整合 AWS Glue Data Catalog	您現在可以將 S3 資料表整合並編目為 AWS Glue Data Catalog 物件，並將目錄註冊為 Lake Formation 資料位置。如需詳細資訊，請參閱《》中的建立 Amazon S3 Tables 目錄 AWS Glue Data Catalog。	2025 年 3 月 13 日
Lake Formation 已更新AWSLakeFormationCrossAccountManager 政策。	Lake Formation 透過將StringLike 條件運算子取代為允許 IAM 執行 ARN 格式檢查的運算ArnLike子，增強了 AWSLakeFormationCrossAccountManager 政策。	2025 年 1 月 25 日
已更新政策變更	記錄 AWSLakeFormationDataAdmin 政策的變更。	2024 年 12 月 3 日
多目錄更新	AWS Glue Data Catalog 可讓您建立聯合目錄、跨 Amazon S3 資料湖和 Amazon Redshift 資料倉儲統一資料，以及整合來自 Amazon DynamoDB 等營運資料庫以及 Snowflake、MySQL 等第三方資料來源	2024 年 12 月 3 日

	的資料。如需詳細資訊，請參閱將資料帶入 AWS Glue Data Catalog。	
更新 LF-Tag 表達式的文件	您可以儲存 LF-Tag 表達式，並重複使用它們來授予 Data Catalog 資源的許可。如需詳細資訊，請參閱管理 LF-Tag 表達式。	2024 年 11 月 7 日
更新 Data Catalog 檢視的文件	除了 AWS Glue Data Catalog 使用 Amazon Athena 和 Amazon Redshift DDLs 之外，您還可以使用 AWS Glue APIs 在 中建立檢視。如需詳細資訊，請參閱建置資料目錄檢視。	2024 年 8 月 7 日
新增可稽核憑證販賣的文件	Lake Formation 可讓您在 CloudTrail 事件中包含 IAM Identity Center 使用者的內容，然後追蹤存取 資源的使用者。如需詳細資訊，請參閱 CloudTrail 日誌中的包含 IAM Identity Center 使用者內容。	2024 年 7 月 14 日
更新政策變更	記錄 AWSLakeFormationCrossAccountManager 和 AWSLakeFormationDataAdmin 政策的變更 （新增陳述式 IDs 和移除的備援許可）。 AWSLakeFormationCrossAccountManager AWSLakeFormationDataAdmin	2024 年 3 月 14 日
更新設定 Lake Formation	已更新設定 AWS Lake Formation區段中的步驟。	2024 年 2 月 7 日

更新政策變更	已將新許可新增至服務連結角色的內嵌政策。如需詳細資訊，請參閱使用 Lake Formation 的服務連結角色。	2024 年 2 月 7 日
已更新政策變更	記錄 LakeFormationDataAccessServiceRolePolicy 政策的變更。	2024 年 2 月 2 日
合併 Lake Formation 限制	針對 Lake Formation 限制和考量事項建立統一區段。如需詳細資訊，請參閱 Lake Formation 限制。	2023 年 12 月 15 日
新增 Iceberg 壓縮的文件	為了獲得 Athena 和 Amazon EMR 等 AWS 分析服務以及 AWS Glue ETL 任務的更佳讀取效能， 為 Data Catalog 中的 Iceberg 資料表 AWS Glue Data Catalog 提供受管壓縮（將小型 Amazon S3 物件壓縮為較大物件的程序）。如需詳細資訊，請參閱最佳化 Iceberg 資料表。	2023 年 11 月 25 日
新增 IAM Identity Center 整合的文件	IAM Identity Center 整合允許使用者和群組存取強制執行 Lake Formation 許可的 Data Catalog 資源。如需詳細資訊，請參閱 IAM Identity Center 整合。	2023 年 11 月 25 日

新增 Data Catalog 檢視的文件	您可以在 中使用 Amazon Athena 或 Amazon Redshift 的 SQL 編輯器 AWS Glue Data Catalog ，在 中建立最多參考 10 個資料表的檢視。如需詳細資訊，請參閱建立檢視。	2023 年 11 月 25 日
已更新政策變更	已記錄 AWSLakeFormationCrossAccountManager 政策的變更。	2023 年 10 月 25 日
新增混合存取模式的文件	混合存取模式可讓您靈活地選擇性地為 中的資料庫和資料表啟用 Lake Formation 許可 AWS Glue Data Catalog。使用混合存取模式時，您現在有一個增量路徑，可讓您為一組特定使用者設定 Lake Formation 許可，而不會中斷其他現有使用者或工作負載的許可政策。如需詳細資訊，請參閱混合存取模式。	2023 年 9 月 26 日
新增建立 Apache Iceberg 資料表的文件	您現在可以建立 Apache Iceberg 資料表，在 中使用 Apache Parquet 資料格式 AWS Glue Data Catalog ，並將資料存放在 Amazon S3 中。如需詳細資訊，請參閱建立 Iceberg 資料表。	2023 年 8 月 16 日

新增跨區域資料存取的文件	Lake Formation 支援跨 AWS 區域查詢 Data Catalog 資料表。您可以使用 Athena、Amazon EMR 從其他區域存取區域中的資料，並在指向來源資料庫和資料表的其他區域中建立資源連結，以執行 AWS Glue ETL。您可以將 Data Catalog 連接到外部中繼存放區，以存放 Amazon S3 資料的中繼資料，並使用 安全地管理資料存取許可 AWS Lake Formation。如需詳細資訊，請參閱跨區域存取資料表。	2023 年 6 月 30 日
重新組織的內容	本指南中的重組章節，以符合 Lake Formation 使用者旅程。	2023 年 5 月 15 日
新增 HMS 聯合的文件	您可以將 Data Catalog 連接到外部中繼存放區，以存放 Amazon S3 資料的中繼資料，並使用 安全地管理資料存取許可 AWS Lake Formation。如需詳細資訊，請參閱管理使用外部中繼存放區的資料集許可。	2023 年 4 月 15 日
新增 Amazon Redshift 資料共用的文件	您現在可以使用 Lake Formation 許可從 Amazon Redshift 安全地管理資料共用中的資料。Lake Formation 支援透過 授權存取您的資料 AWS Data Exchange。如需詳細資訊，請參閱 中的資料共用 AWS Lake Formation。	2022 年 11 月 30 日

支援直接與委託人共用跨帳戶資料	新增直接與另一個帳戶中的 IAM 主體共用資料的相關資訊。如需詳細資訊，請參閱 中的跨帳戶資料共用 AWS Lake Formation。	2022 年 11 月 10 日
支援使用 TBAC AWS RAM 啟用的資料共用	新增有關將 Data Catalog 許可 AWS Resource Access Manager 使用授予跨帳戶授予的 LF-TBAC 方法的資訊。	2022 年 11 月 10 日
新增使用其他 服務的章節	新增有關 Athena、 AWS Glue Redshift Spectrum 和 Amazon EMR 等 AWS 服務如何使用 Lake Formation 安全地存取向 Lake Formation 註冊之 Amazon S3 位置中的資料的資訊。如需更多資訊，請參閱使用其他 AWS 服務。	2022 年 11 月 10 日
???	新增使用 Amazon EMR 存取跨帳戶資料時疑難排解錯誤的相關資訊。如需詳細資訊，請參閱使用 Amazon EMR 存取跨帳戶共用的資料時發生錯誤。	2022 年 11 月 7 日
跨帳戶資源共享的更新	新增描述，說明跨帳戶資源共用如何在 Lake Formation 中運作。記錄 AWSLakeFormationCrossAccountManager 政策的變更。	2022 年 5 月 6 日
新的教學課程	新增了有關建立受管資料表、保護資料湖和共用資料湖的新教學課程。如需詳細資訊，請參閱入門一節。	2022 年 4 月 20 日

新 Lake Formation 登陸頁面	更新 Lake Formation 登陸頁面，納入教學課程的連結，提供如何使用 Lake Formation 建置資料湖、擷取資料、共用和保護資料湖的step-by-step說明。	2022 年 4 月 20 日
憑證販賣的支援	新增憑證販賣的相關資訊，其支援 Lake Formation，允許第三方服務使用憑證販賣 API 操作與 Lake Formation 整合。如需詳細資訊，請參閱登入資料販賣在 Lake Formation 中的運作方式。	2022 年 2 月 28 日
支援受管資料表和進階資料篩選	新增受管資料表的相關資訊，以支援 ACID 交易、自動資料壓縮和時間歷程查詢。新增有關建立資料篩選條件以支援資料欄層級安全性、資料列層級安全性和儲存格層級安全性的資訊。如需詳細資訊，請參閱 Lake Formation 中的受管資料表和 Lake Formation 中的資料篩選和儲存格層級安全性。	2021 年 11 月 30 日
支援 VPC 介面端點	新增有關為 Lake Formation 建立虛擬私有雲端 (VPC) 介面端點的資訊，以便 VPC 和 Lake Formation 之間的通訊完全安全地在 AWS 網路中執行。如需詳細資訊，請參閱搭配使用 Lake Formation 與 VPC 端點。	2021 年 10 月 11 日

支援 VPC 端點政策	新增支援 Lake Formation 中虛擬私有雲端 (VPC) 端點政策的相關資訊。如需詳細資訊，請參閱搭配使用 Lake Formation 與 VPC 端點。	2021 年 10 月 11 日
支援標籤型存取控制	Lake Formation 標籤型存取控制提供了一種新的、更具可擴展性的方法，透過使用 LF 標籤來管理對 Data Catalog 資源和基礎資料的存取。如需詳細資訊，請參閱 Lake Formation 標籤型存取控制。	2021 年 5 月 7 日
Amazon EMR 上資料篩選的新選擇加入要求。	新增選擇加入以允許 Amazon EMR 篩選 Lake Formation 管理之資料的相關資訊。如需詳細資訊，請參閱允許 Amazon EMR 上的資料篩選。	2020 年 10 月 9 日
支援授予 Data Catalog 資料庫的完整跨帳戶許可	新增有關跨 AWS 帳戶對 Data Catalog 資料庫授予完整 Lake Formation 許可的資訊，包括 `CREATE_TABLE` 。如需詳細資訊，請參閱共用資料目錄資料庫。	2020 年 10 月 1 日
支援透過 SAML 驗證 Amazon Athena 使用者。	新增支援透過 JDBC 或 ODBC 驅動程式連線並透過 Okta 和 Microsoft Active Directory Federation Service (AD FS) 等 SAML 身分提供者進行身分驗證的 Athena 使用者的相關資訊。如需詳細資訊，請參閱AWS 服務整合與 Lake Formation。	2020 年 9 月 30 日

支援使用加密的資料目錄進行跨帳戶存取	新增在 Data Catalog 加密時授予跨帳戶許可的相關資訊。如需詳細資訊，請參閱跨帳戶存取先決條件。	2020 年 7 月 30 日
支援跨帳戶存取資料湖	新增有關將 Data Catalog 資料庫和資料表的 AWS Lake Formation 許可授予外部 AWS 帳戶和組織的資訊，以及有關存取從外部帳戶共用的 Data Catalog 物件的資訊。如需詳細資訊，請參閱跨帳戶存取。	2020 年 7 月 7 日
與 Amazon QuickSight 整合	新增如何將 Lake Formation 許可授予 QuickSight Enterprise Edition 使用者的相關資訊，以便他們可以存取位於已註冊 Amazon S3 位置的資料集。如需詳細資訊，請參閱授予資料目錄許可。	2020 年 6 月 29 日
設定和入門章節的更新	重組並改善設定和入門章節。更新資料湖管理員的 recommended AWS Identity and Access Management (IAM) 許可。	2020 年 2 月 27 日

的支援 AWS Key Management Service	新增 Lake Formation 如何支援 AWS Key Management Service (AWS KMS) 的資訊，簡化了在已註冊的 Amazon Simple Storage Service (Amazon S3) 位置中讀取和寫入加密資料的整合服務設定。新增如何註冊加密的 Amazon S3 位置的相關資訊 AWS KMS keys。如需詳細資訊，請參閱the section called “將 Amazon S3 位置新增至您的資料湖”。	2020 年 2 月 27 日
藍圖和資料湖管理員 IAM 政策的更新	釐清增量資料庫藍圖的輸入參數。更新資料湖管理員所需的 IAM 政策。	2019 年 12 月 20 日
安全章節重寫和升級章節修訂	已改善安全性和升級章節。	2019 年 10 月 29 日
超級許可取代所有許可	更新安全性和升級章節，以反映`All`以 取代許可`Super`。	2019 年 10 月 10 日
新增、更正和釐清	根據意見回饋進行新增、更正和釐清。已修訂安全章節。更新安全性和升級章節，以反映`Everyone`以 取代 群組`IAMAllowedPrincipals` 。	2019 年 9 月 11 日
新的指南	這是初版的 AWS Lake Formation 開發人員指南。	2019 年 8 月 8 日

AWS 詞彙表

如需最新的 AWS 術語，請參閱 AWS 詞彙表 參考中的AWS 詞彙表。

本文為英文版的機器翻譯版本，如內容有任何歧義或不一致之處，概以英文版為準。