AWS BLACK BELT ONLINE SEMINAR

AWS Lake Formation

基礎編

佐藤 祥多 (Shota Sato)

Analytics Specialist Solution Architect

AWS Black Belt Online Seminar とは

- 「サービス別」「ソリューション別」「業種別」などのテーマに分け、アマゾン ウェブ サービス ジャパン合同会社が提供するオンラインセミナーシリーズです
- AWS の技術担当者が、AWS の各サービスやソリューションについてテーマごとに動画を公開します
- 動画を一時停止・スキップすることで、興味がある分野・項目だけの聴講も可能、スキマ時間の学習にもお役立ていただけます
- 以下の URL より、過去のセミナー含めた資料などをダウンロードすることができます
  - https://aws.amazon.com/jp/aws-jp-introduction/aws-jp-webinar-service-cut/

内容についての注意点

- 本資料では 2023 年 09 月時点のサービス内容および価格についてご説明しています。最新の情報は AWS 公式ウェブサイト (https://aws.amazon.com/) にてご確認ください
- 資料作成には十分注意しておりますが、資料内の価格と AWS 公式ウェブサイト記載の価格に相違があった場合、AWS 公式ウェブサイトの価格を優先とさせていただきます
- 価格は税抜表記となっています。日本居住者のお客様には別途消費税をご請求させていただきます

自己紹介

名前：**佐藤 祥多 (Shota Sato)**

所属：アマゾンウェブサービスジャパン
　アナリティクス事業本部
　ソリューションアーキテクト本部
　アナリティクスソリューションアーキテクト

好きなAWSサービス：
AWS Lake Formation, Amazon DataZone, AWS Glue

アジェンダ

- AWS Lake Formation の概要
- AWS Lake Formation の機能
  - アクセス制御の概要と機能
  - ブループリントの概要と機能
  - データ共有の概要と機能
- AWS Lake Formation の料金
- まとめ

AWS Lake Formation の概要

AWS Lake Formation

安全なデータレイクを簡単に作成し
幅広い分析にデータを利用可能にする

使い慣れたデータベースのような機能を使用して、
データレイクをすばやく作成、管理、保護

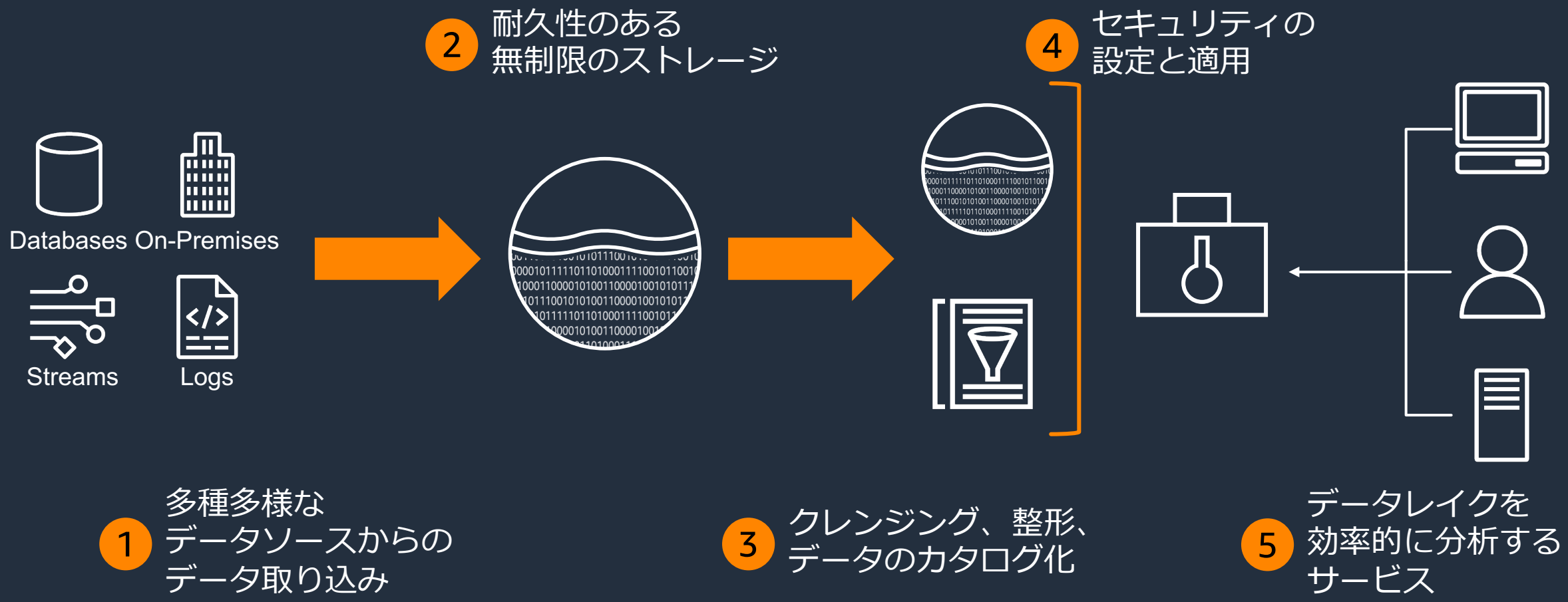
大規模なセキュリティ管理とガバナンスを簡素化し、
データレイク全体できめ細かなアクセス許可が可能

一元化されたデータカタログで
すべてのデータを検出、アクセス

クロスアカウントデータ共有により、
組織全体のデータアクセスをサポート

データレイクを作るためには

数多くの考慮事項があり、それぞれ適切な対応が必要

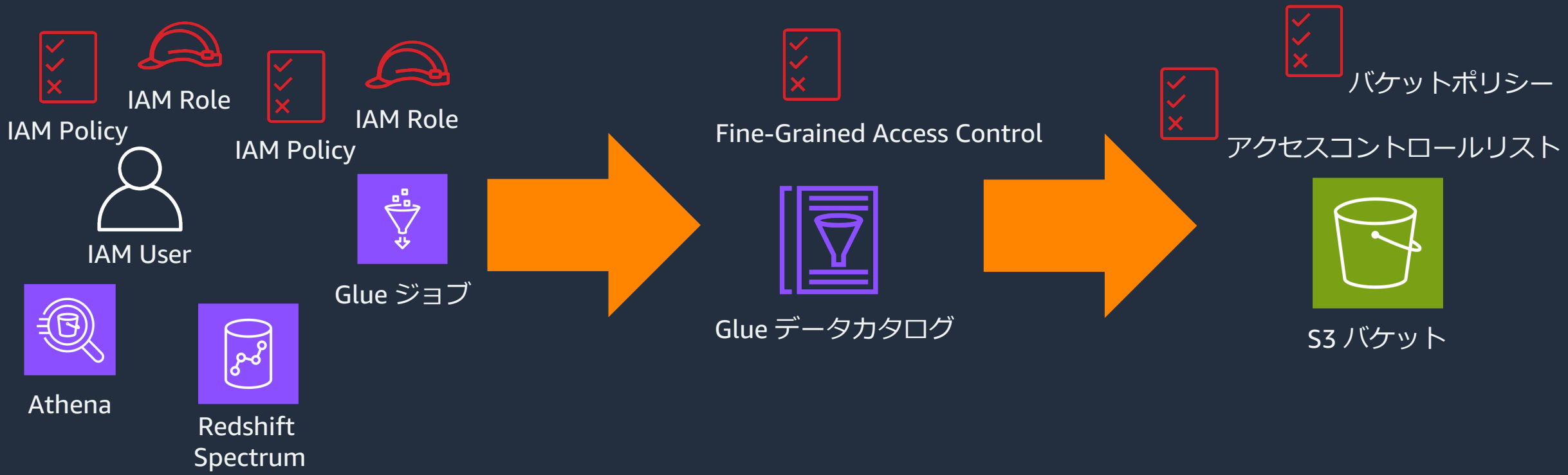


安全なデータレイクを作るための AWS のソリューション

それぞれの考慮事項に対して AWS では様々なソリューションを提供

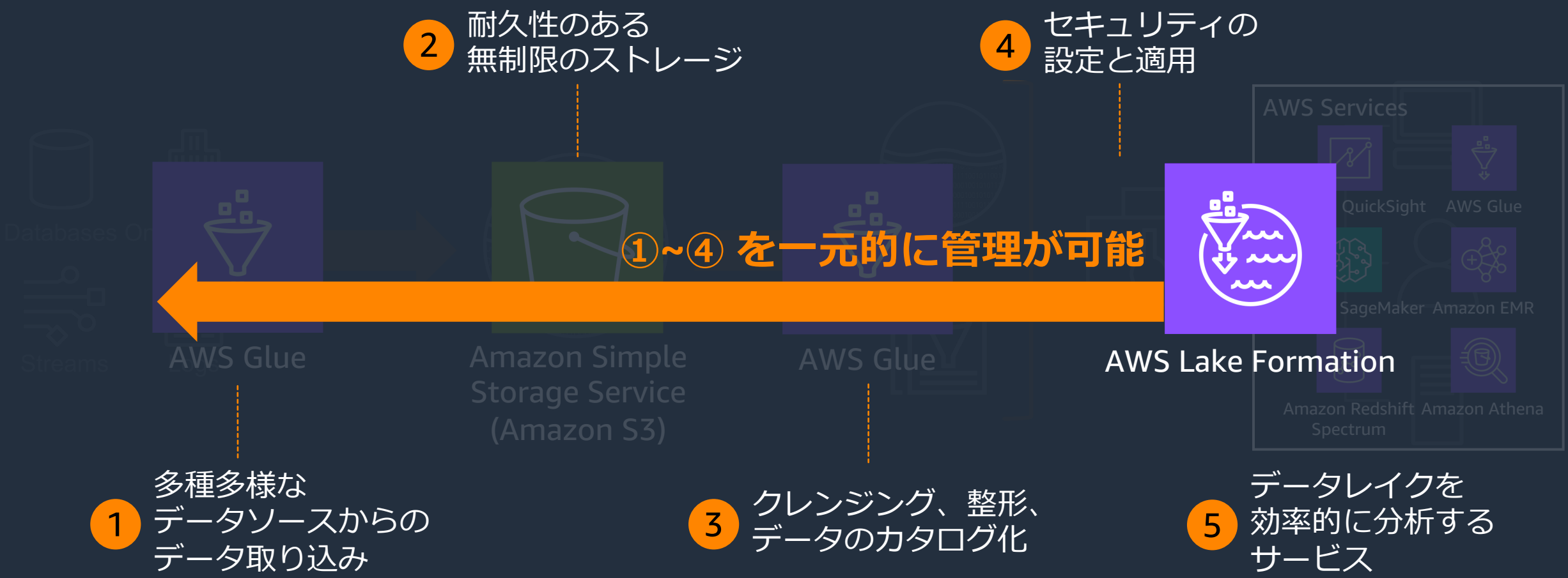
従来のデータレイク認可の課題

IAMポリシーとバケットポリシーなどを複数の異なる箇所にポリシーを登録する必要があり、
管理が煩雑だった



AWS Lake Formation の導入効果

データレイクへのデータ取り込みからセキュリティの設定/適用を一元的に実施可能



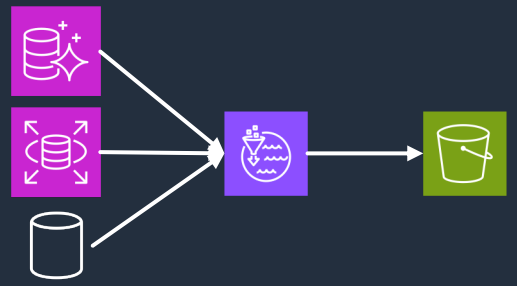
AWS Lake Formation の主要な機能

安全なデータレイクを簡単に作成するために Lake Formation には3つの主要機能がある

一元管理による
細かな粒度のアクセス制御

データレイク管理者が AWS サービス の S3 へのアクセスを一元的に制御

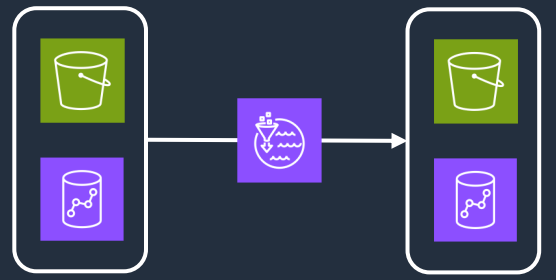
細かなアクセス制御に基づいて行レベル/列レベル/セルレベルのセキュリティを実現



ブループリント機能を利用したデータの取り込み

Glue の Connection を利用して様々なデータソースからデータレイクにデータ取り込み

洗い替えだけでなく増分的なデータ取り込みにも対応

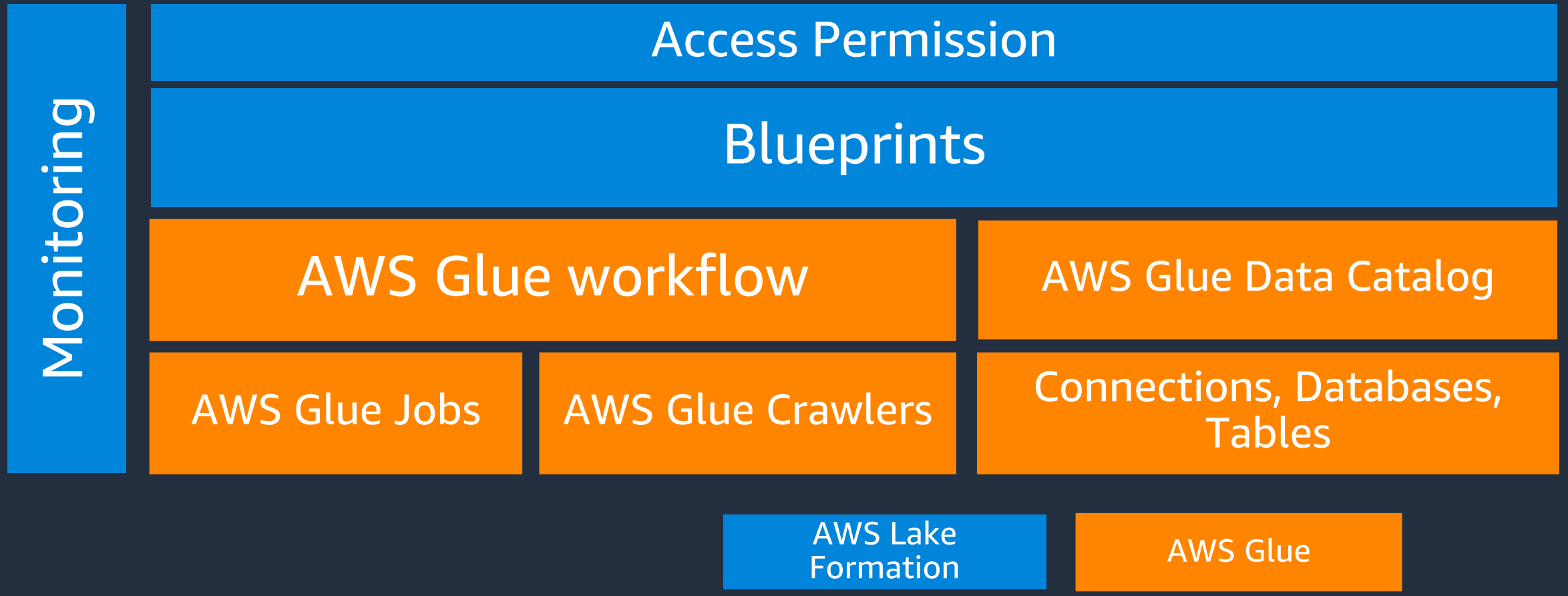


安全なデータ共有の制御

S3 のデータレイクや Redshift のデータを安全に別 AWS アカウントや別リージョンに共有し制御可能

補足：AWS Lake Formation と AWS Glue の関係

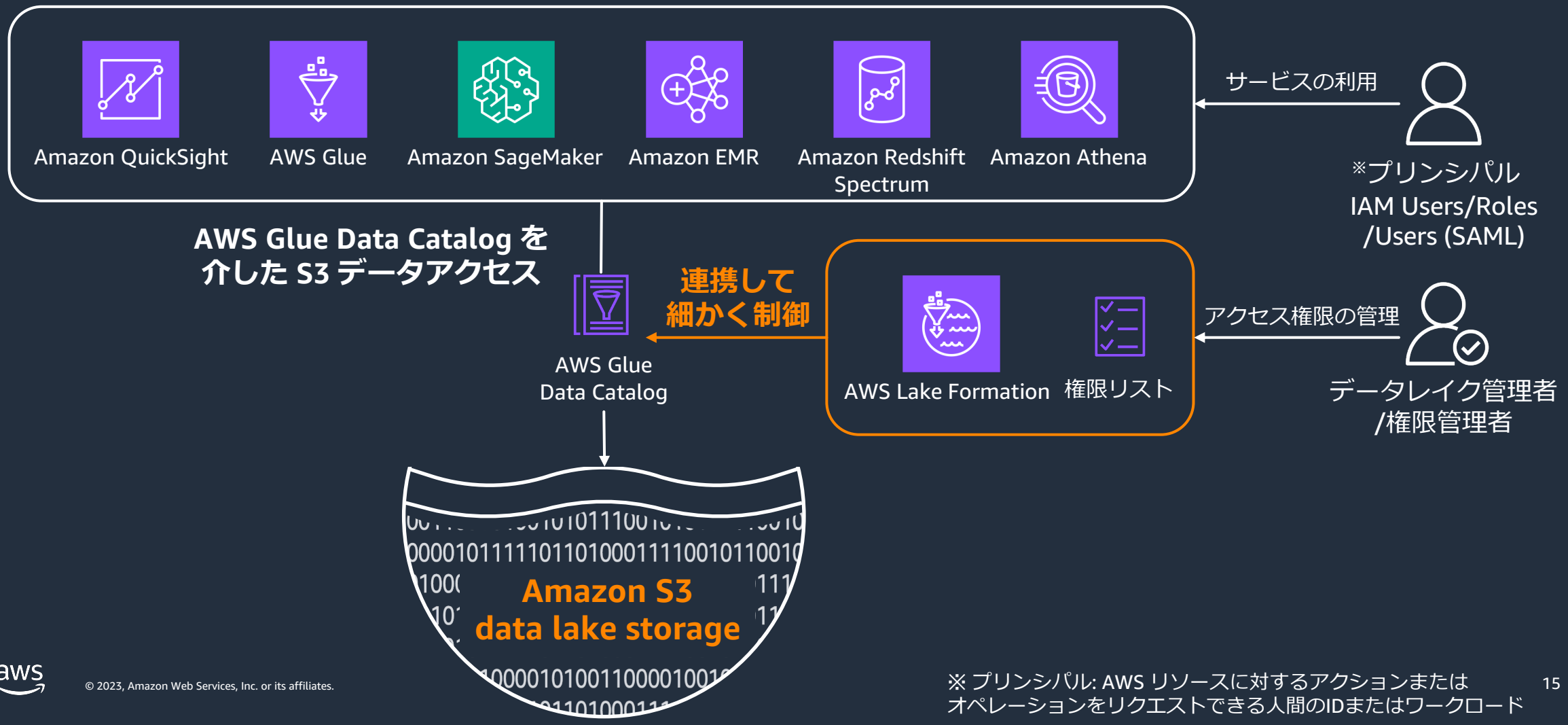
AWS Lake Formation は AWS Glue の拡張機能と言え、Glue Data Catalog へのセキュリティ強化やブループリントによるデータ取り込み等、データレイク構築に必要な機能を一元的に提供



AWS Lake Formation の機能 - アクセス制御の概要と機能

AWS Lake Formation のアクセス制御の全体像

AWS Lake Formation は AWS Glue Data Catalog を介した Amazon Simple Storage Service (S3) へのアクセスを細かな粒度で制御する



※ プリンシパル: AWS リソースに対するアクションまたはオペレーションをリクエストできる人間のIDまたはワークロード

AWS Lake Formation を利用する際のユーザロール

アクセス制御を管理するためにはユーザのロールを整理すると良い。Lake Formation の利用では3つのロールに分類できる

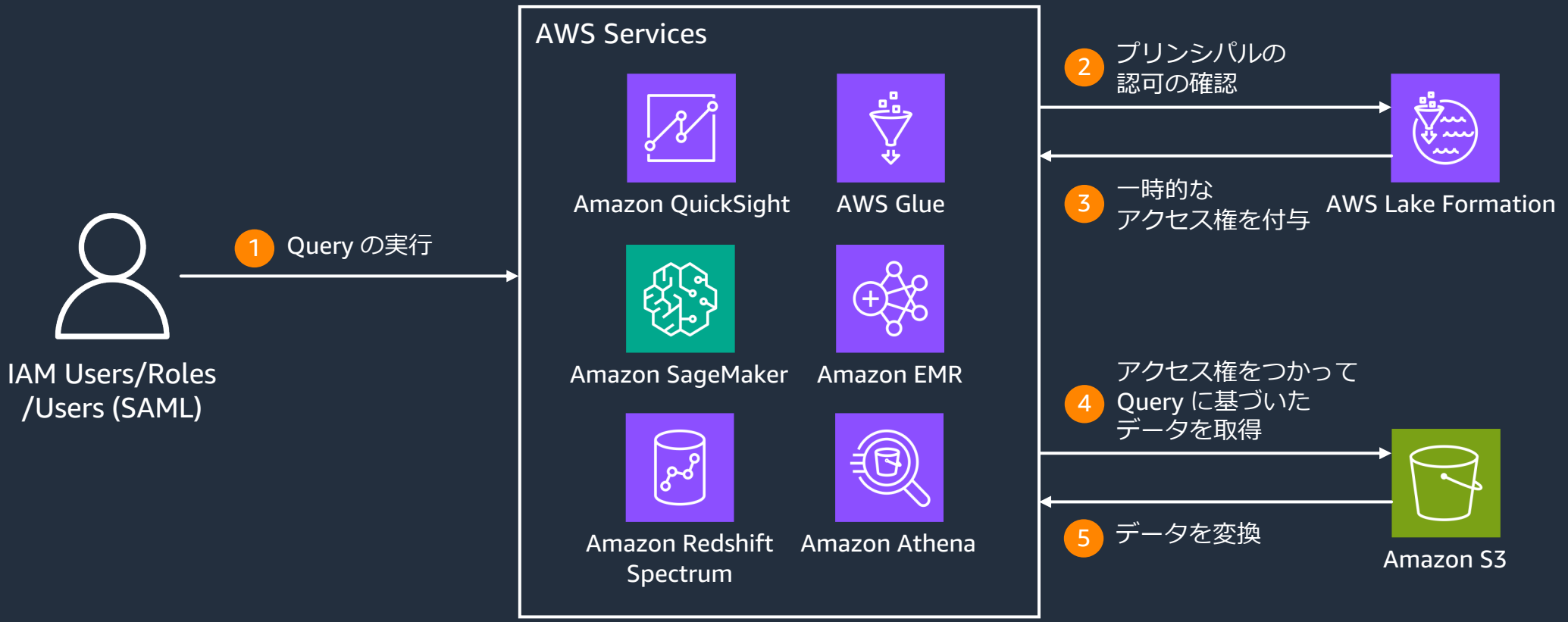
- データレイクの構築/管理を行う
- あらゆるリソースやデータロケーションへのアクセス許可を他のプリンシパルに付与可能
- アクセス許可の付与を行う許可を他のプリンシパルに付与可能で権限管理者に権限付与権限を付与する

- アクセス権限を管理する
- データレイク管理者から付与された権限付与権限を利用して利用者へアクセス権限を付与する

- データレイクを利用する

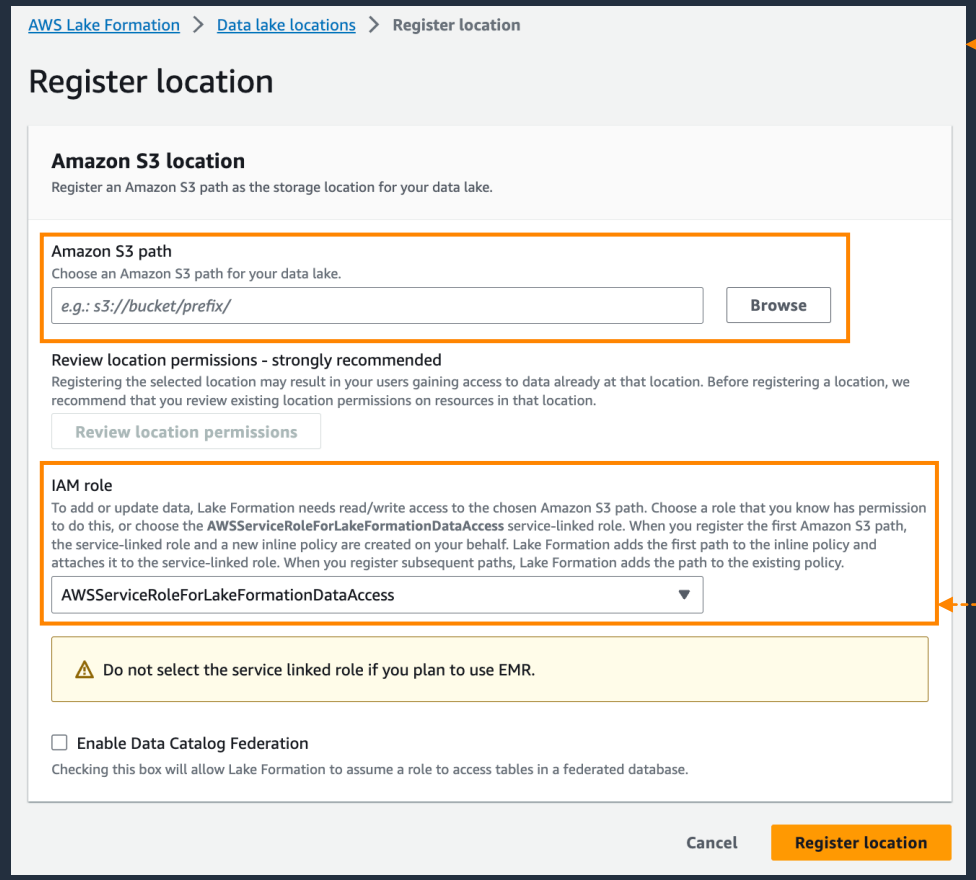
アクセス制御の裏側

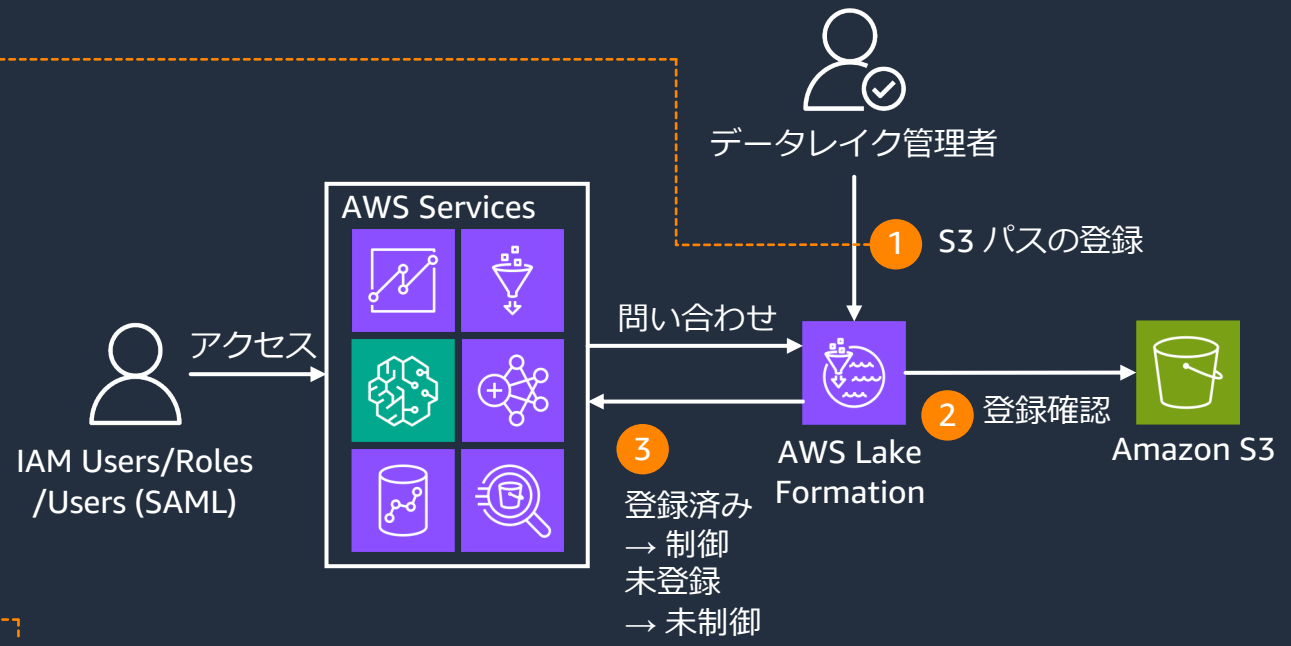
AWS Glue Data Catalog を介した S3 へのアクセスが発生すると、AWS Lake Formation は AWS サービスに一時的な認証情報を返して、サービスはその一時的な認証情報をつかって S3 にアクセスする



アクセス制御の対象となる S3 パスの登録

AWS Lake Formation は事前に登録された S3 パスの配下にあるリソースのみアクセス制御を行うため、事前に Glue Data Catalog のリソースがポイントする S3 パスを登録する必要がある

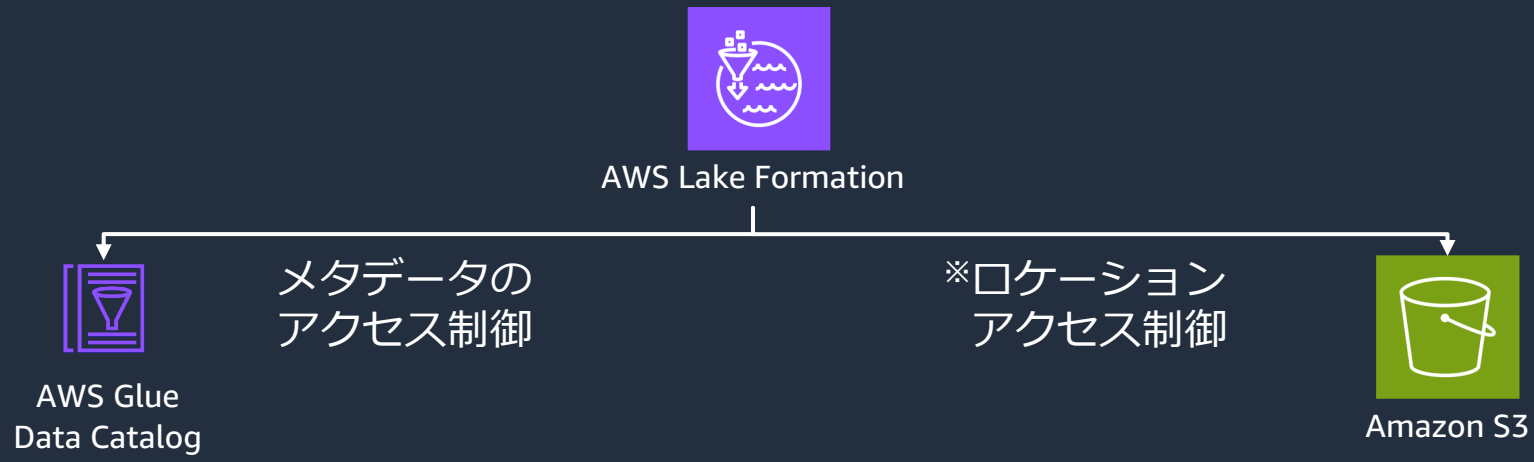




- Lake Formation が S3に write/read を行うための IAM ロール を指定する
- AWSServiceRoleForLakeFormationDataAccess がデフォルトで指定されてる

AWS Lake Formation のアクセス制御対象

AWS Lake Formation は AWS Glue Data Catalog に対するメタデータのアクセス制御と
Amazon S3 ロケーション内のデータへのアクセス制御を行っている



- **プリンシパルが Glue Data Catalog のメタデータにアクセスする権限を制御**
- [Data lake permissions] から設定可能
- IAM Principals/SAML/External Accounts に Glue Data Catalog の DB/Table への *CRUD* + 参照 (Describe) 権限を付与

- **プリンシパルが S3 をポインタする Glue Data Catalog のメタデータを作成/変更する権限を制御**
- [Data locations] から設定可能
- IAM Principals/SAML/External Accounts に Glue Data Catalog がポイントするS3 リソースへの *CREATE_TABLE/ALTER* 権限を付与

※ 前述された S3 パスの登録とロケーションアクセス制御は別物。S3 パスの登録は Lake Formation が管理するパスを登録し、ロケーションアクセス制御は IAM Principals が S3 ロケーションにアクセスできる権限を制御

各権限で必要なアクセス制御

各権限で必要なアクセス制御は以下の通り

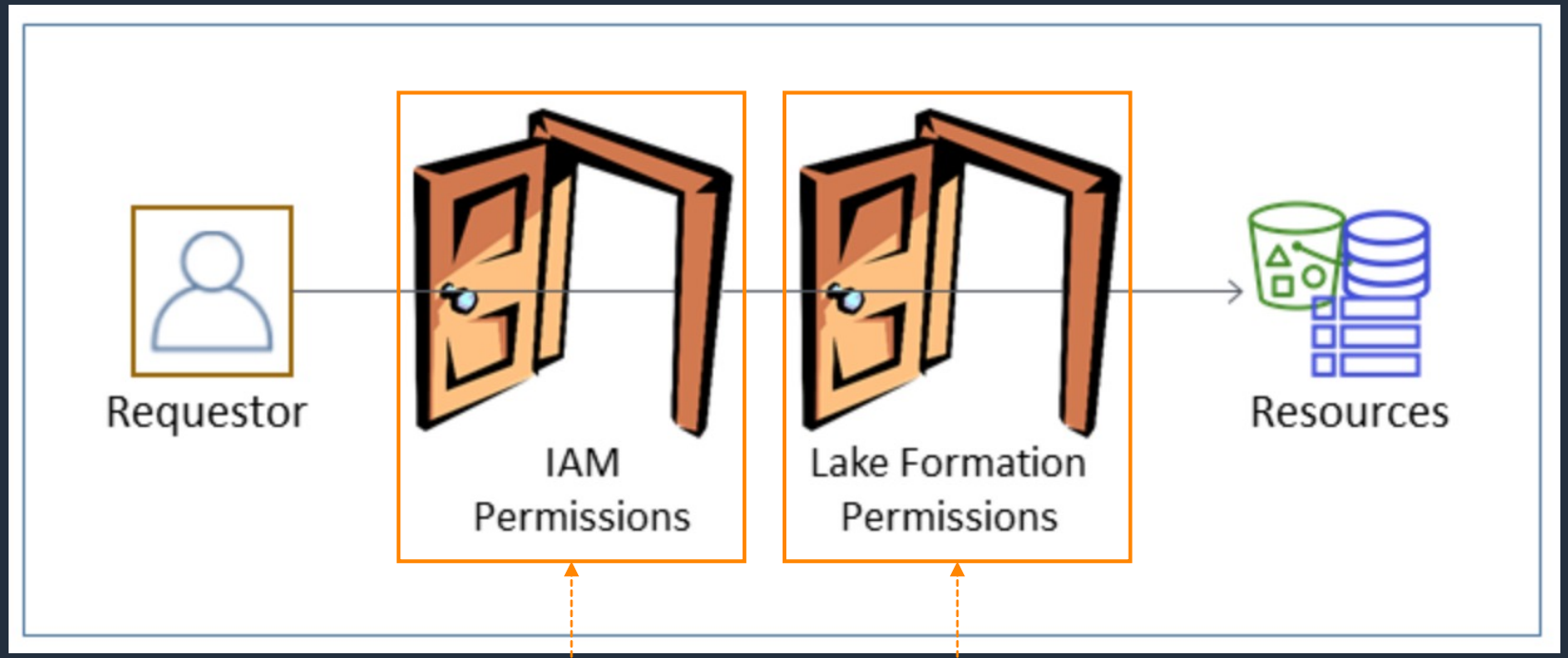
種別	権限	メタデータのアクセス権の要否	ロケーションのアクセス権の要否
DB	Create Table	○	○
	Alter	○	○
	Drop	○	×
	Describe	○	×
Table	Select	○	×
	Insert	○	×
	Delete	○	×
	Describe	○	×
	Alter	○	○
	Drop	○	×

Create Table/Alter はメタデータのアクセス権だけでなくロケーションのアクセス権が必要

※ データレイク管理者はアクセス制御対象となる S3 パスに対するロケーションアクセス権をデフォルトで所持してる

補足：IAM と AWS Lake Formation の制御の関係

AWS Lake Formation によるアクセスの制御は IAM による制御の後に行われる。そのため Lake Formation による細かな制御を行うには、まず IAM による制御を荒くする必要がある



以下のような荒い粒度の制御を IAM では行う
- “athena: *”: 利用する AWS サービスの権限
- “glue:*”: Glue Data Catalog の操作権限
- “lakeformation:*” Lake Formation の操作権限

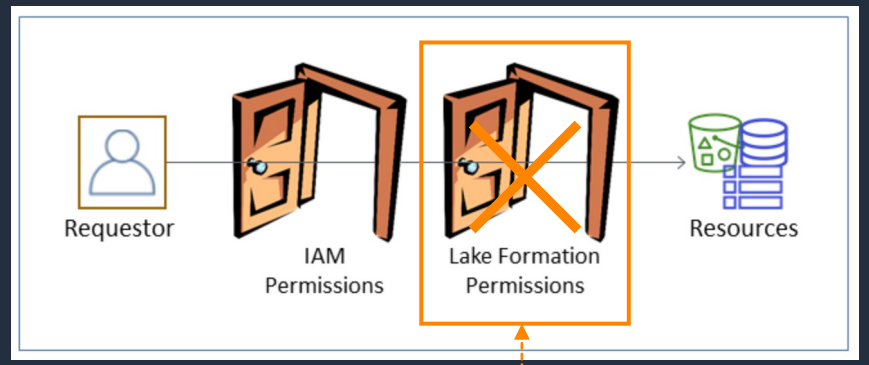
Lake Formation 上で Glue Data Catalog へ細かなアクセス制御を行う

https://docs.aws.amazon.com/lake-formation/latest/dg/lf-permissions-overview.html

補足：IAMAllowedPrincipals

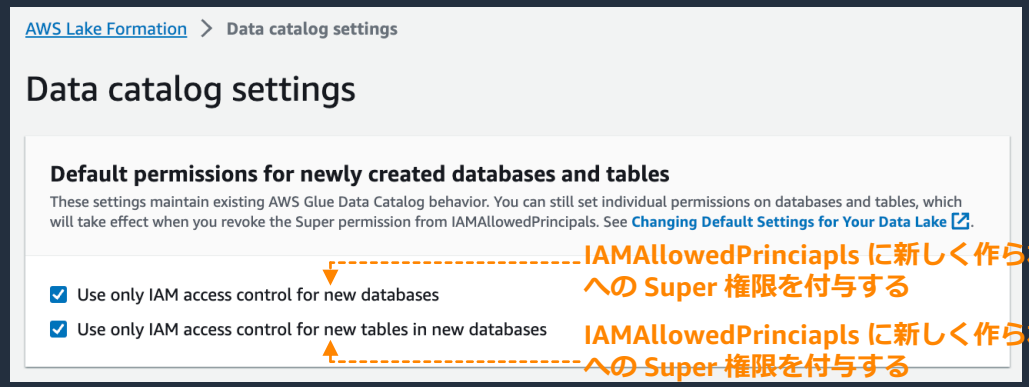
AWS Lake Formation は IAM による制御を妨げないような設定がデフォルトで行われており、
それが IAMAllowedPrincipals による設定である

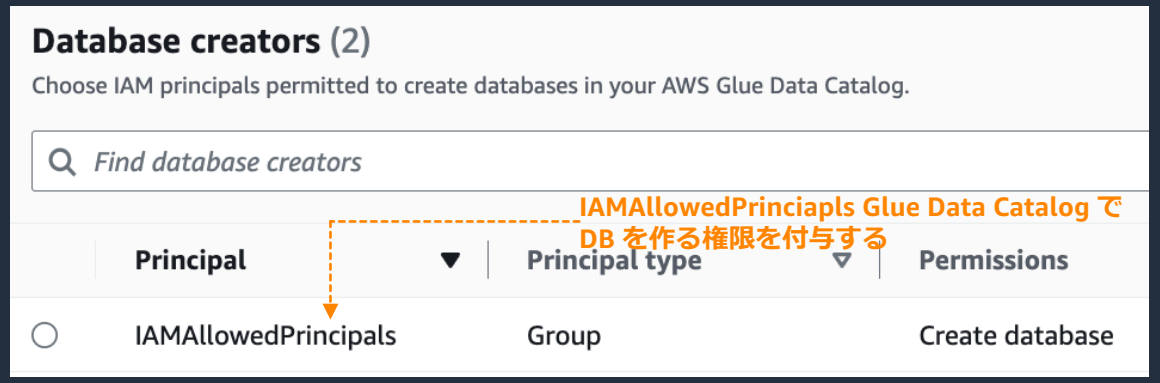
IAMAllowedPrincipals



- ここを無効化する仕組みが IAM Group: IAMAllowedPrincipals による制御
- **IAMAllowedPrincipals には IAM Policy で Glue Data Catalog へのアクセスが許可されてる IAM User/Role が全て含まれる**
- Lake Formation はデフォルトで IAMAllowedPrincipals に Glue Data Catalog の DB/Table 両方に Super 権限を付与する

Lake Formation の初期設定





アクセス制御を始めるためには

※ 設定ドキュメントのリンク　※ 設定ドキュメントのリンク

AWS Glue Data Catalog にリソースがない場合とある場合で開始方法が異なる。
ある場合は IAMAllowedPrincipals の設定をリソースから排除する必要があるが、ない場合含めて基本的な流れは次の通り

1. データレイク管理者用の IAM User/Role の設定
    - Lake Formation を管理するのに必要な IAM Policy の設定をする
2. デフォルトの許可モデルを変更
    - IAMAllowedPrincipals の設定を排除する
3. データレイク用の Amazon S3 ロケーションを設定
    - S3 パスを Lake Formation に登録する
4. データロケーションのアクセス制御を設定
    - S3 へのロケーションアクセス制御の設定をする
5. メタデータのアクセス制御を設定
    - Glue Data Catalog のメタデータへのアクセス制御の設定をする

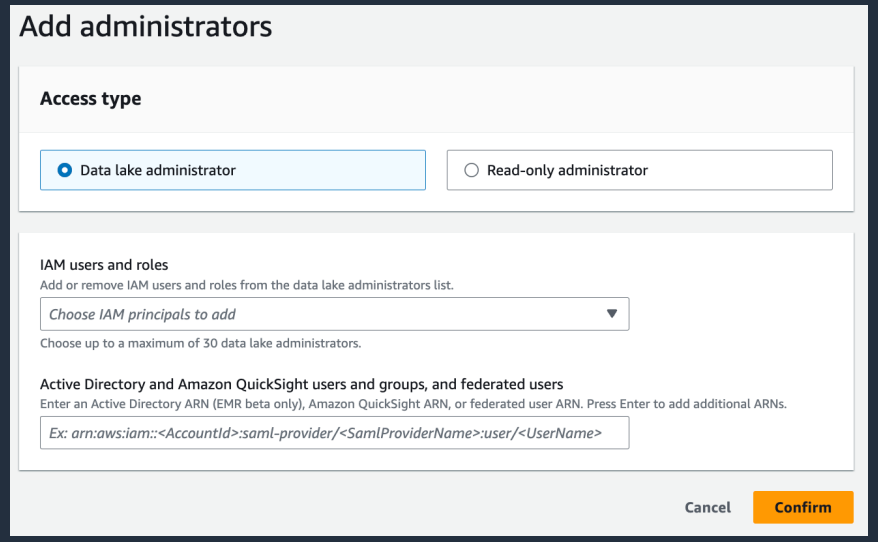
1. データレイク管理者用の IAM User/Role の設定

データレイク管理者は S3 ロケーションの登録と Lake Formation 上のでアクセス制御を管理する。IAM に然るべき Policy を付与した後、Lake Formation 上のコンソールで設定をする

1. 必要な IAM (マネージド) ポリシー

- 必須
  - AWSLakeFormationDataAdmin
  - サービスリンク用インラインポリシー
- オプション
  - AWSGlueConsoleFullAccess
    CloudWatchLogsReadOnlyAccess
    - Glue を使う場合
  - AWSLakeFormationCrossAccountManger
    - クロスアカウント共有を使う場合
  - AmazonAthenaFullAccess
    - Athena を使う場合

2. データレイク管理者の登録

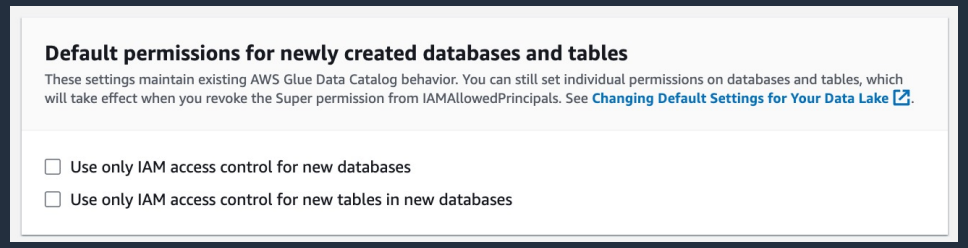


- [Administrative roles and tasks] → [Data lake administrators] → [Add]
- IAM User/Role, SAML ユーザを登録する

2. デフォルトの許可モデルを変更

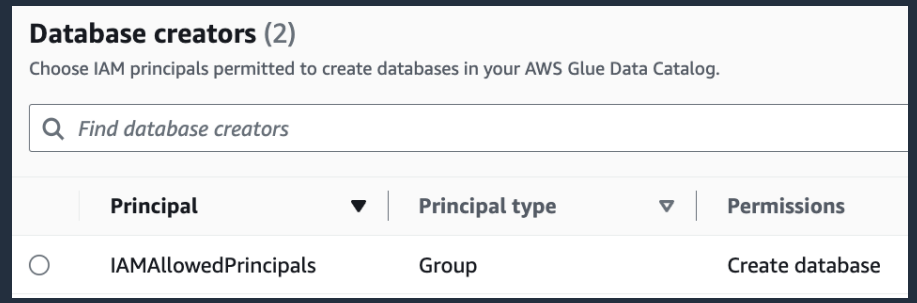
IAMAllowedPrincipals の影響を排除する

1. Data Catalog の設定を変更



- [Data catalog settings]
- チェックボックスをすべて外す
  - 外すと新しく作られる Glue Data Catalog の DB/Table で IAMAllowedPrincipals の設定が追加されなくなる

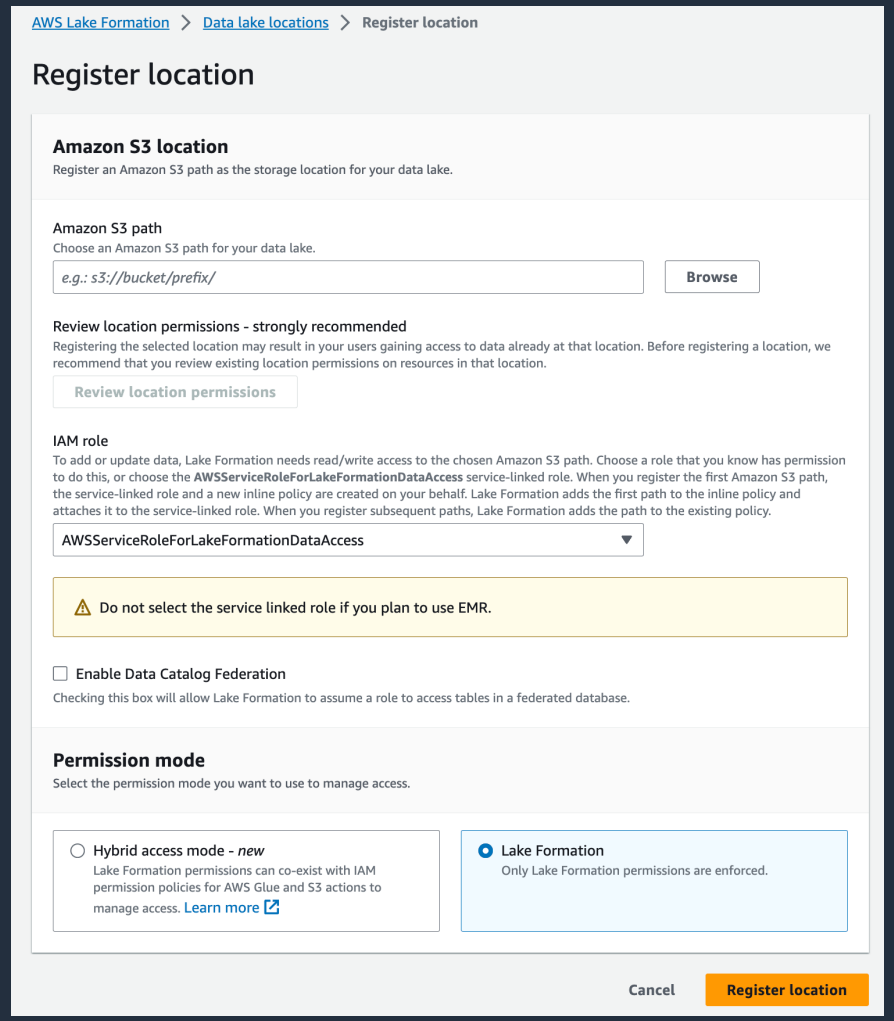
2. IAMAllowdPrincipals の Revoke



- [Administrative roles and tasks] → [Database creators] → [Revoke]
- IAMAllowedPrincipals から DB の作成権限を削除する

3. データレイク用の Amazon S3 ロケーションを設定する

Lake Formation で管理する S3 ロケーションと登録する

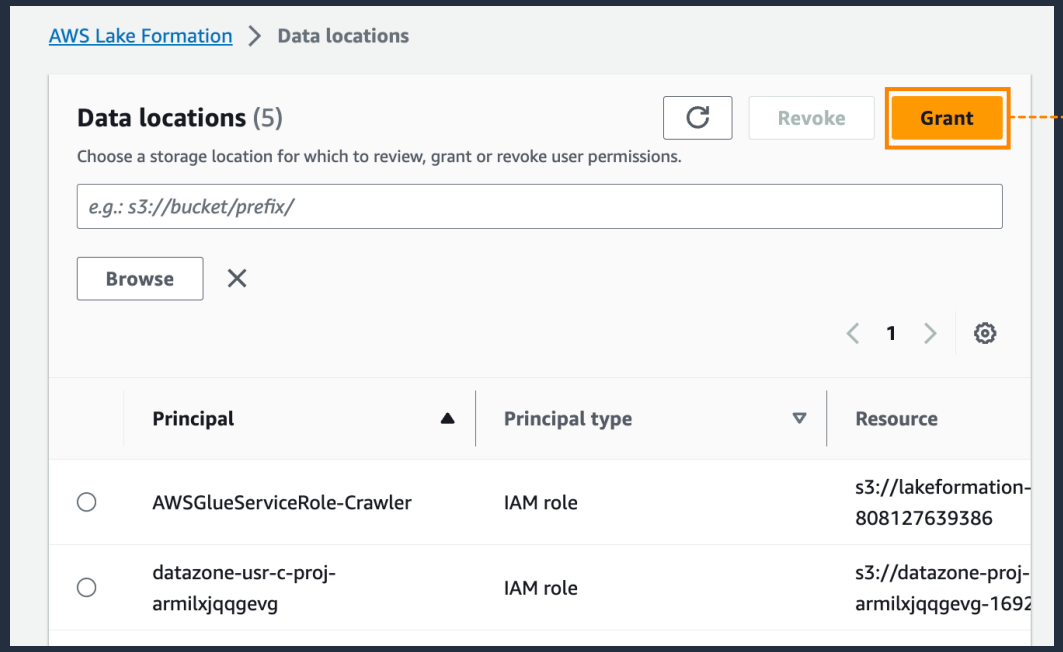


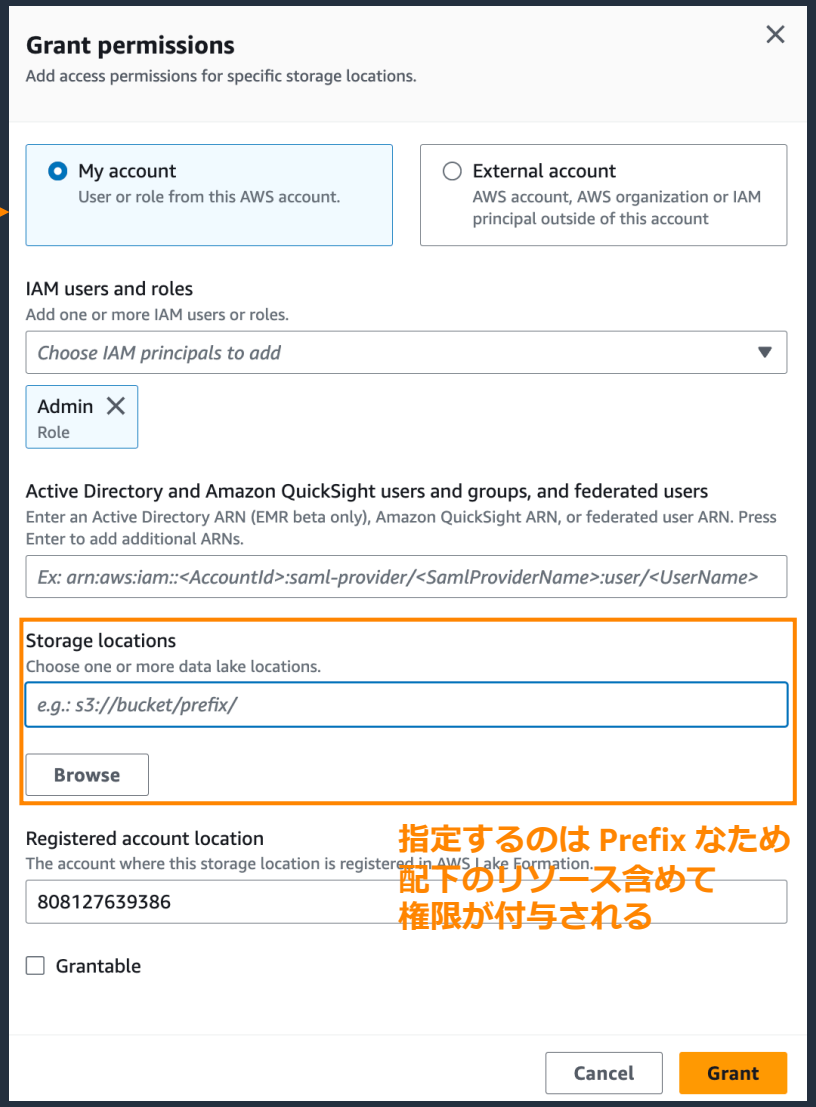
- [Data lake locations]
  → [Register location]
- デフォルトのサービスリンクロールを IAM Role として指定する場合は p16 で作ったインラインポリシーがデータレイク管理者に必要

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

4. データロケーションのアクセス制御を設定

[Data locations] → [Grant] でプリンシパルに権限を付与する

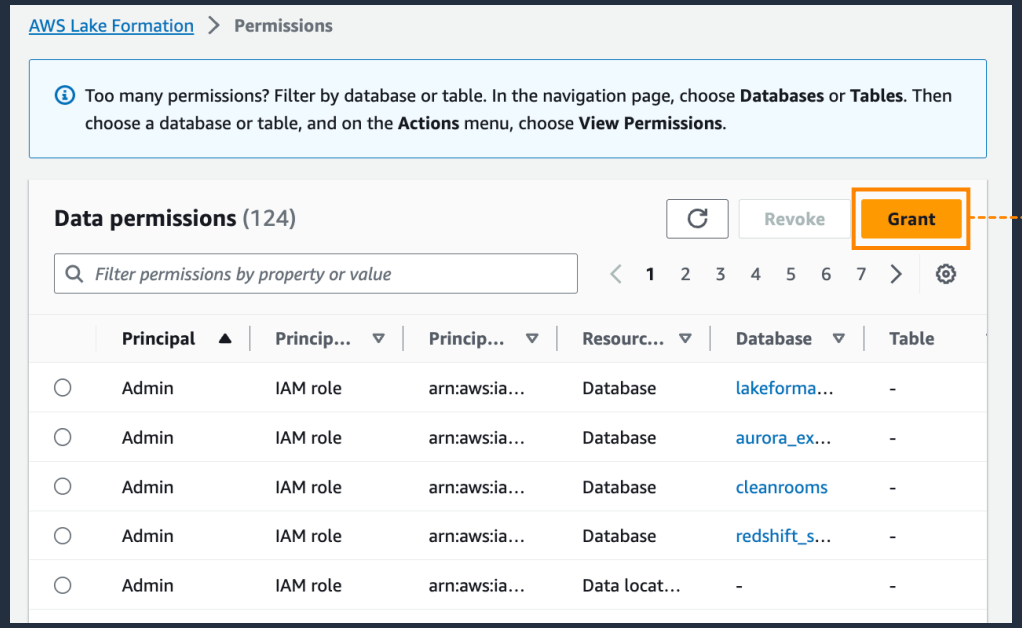


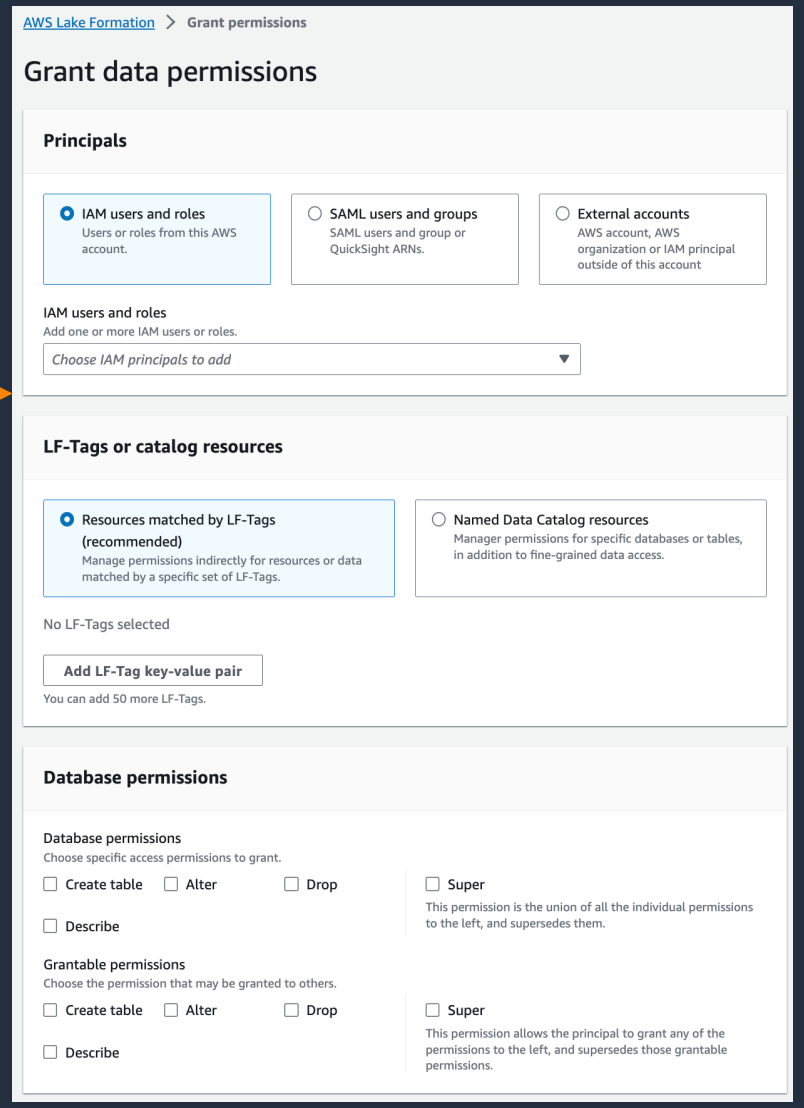


指定するのは Prefix なため
配下のリソース含めて
権限が付与される

5. メタデータのアクセス制御を設定

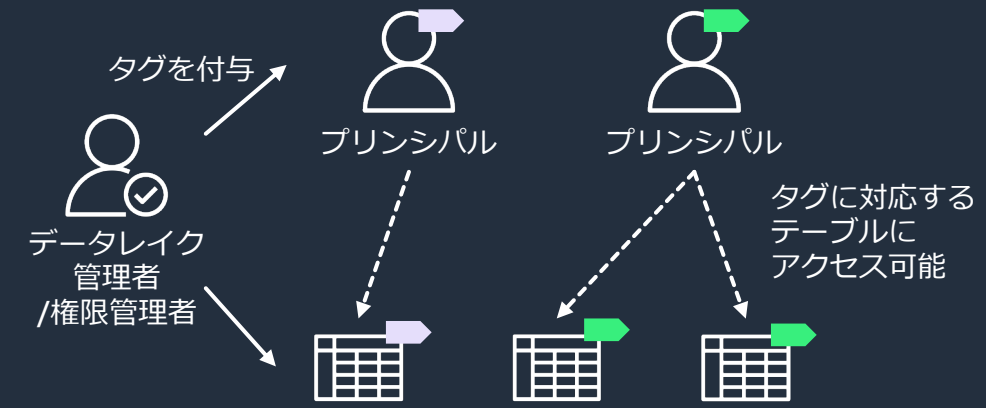
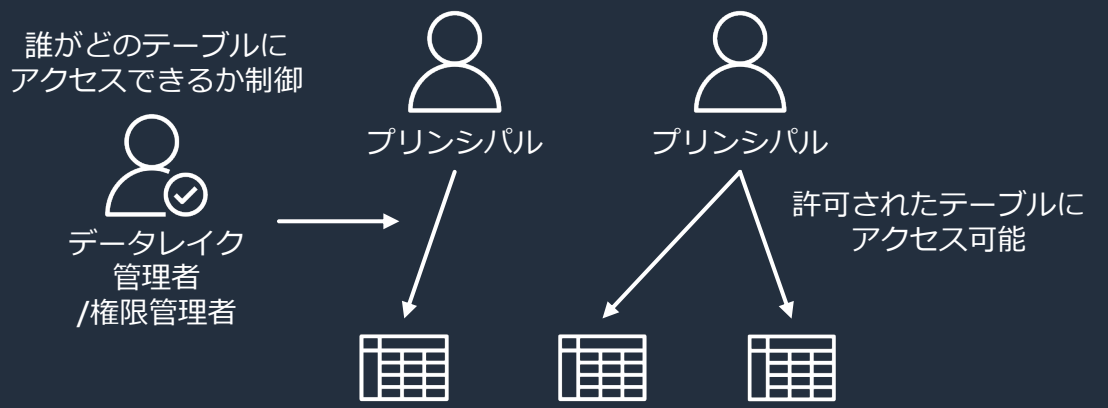
[Data lake permissions] → [Grant] でプリンシパルに権限を付与する





メタデータアクセス制御方法の種類

メタデータのアクセス制御方法には LF Tag を使った方法とリソースを指定する方法の2種があり、それぞれ特徴がある

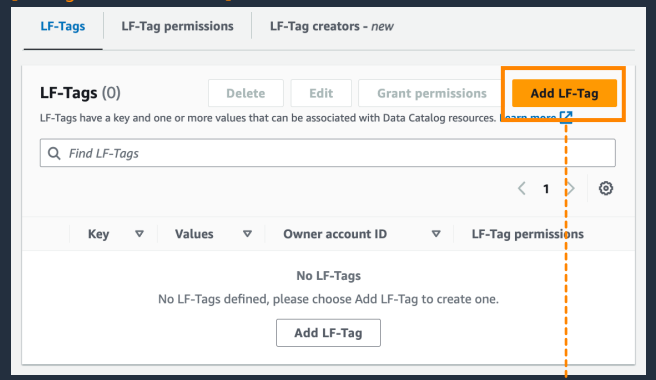
	LF タグ 方式 (推奨)	リソース方式
		
権限制御の数	プリンシパルの数 + テーブルの数	プリンシパルの数 × テーブルの数
セキュリティの細かさ	• LF タグ：列レベル	• シンプルな列フィルタリング：列レベル • データフィルタ：行/列/セルレベル
	利用者とテーブルにそれぞれ タグを付与するためスケールがしやすいが セキュリティは列レベルまで	利用者とテーブルの組み合わせに許可を 付与するためスケールがしにくいが セキュリティは細かく制御可能

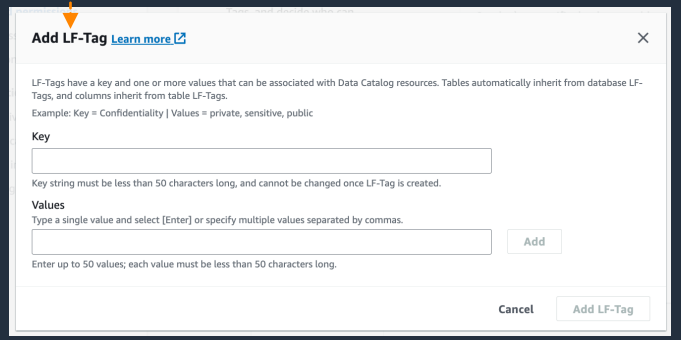
LF タグ方式の設定方法

3つのステップで設定をおこなう

1. LF タグ を定義

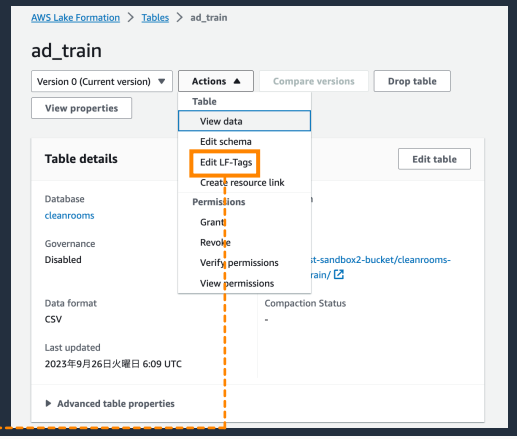
[LF-Tags and Permissions]

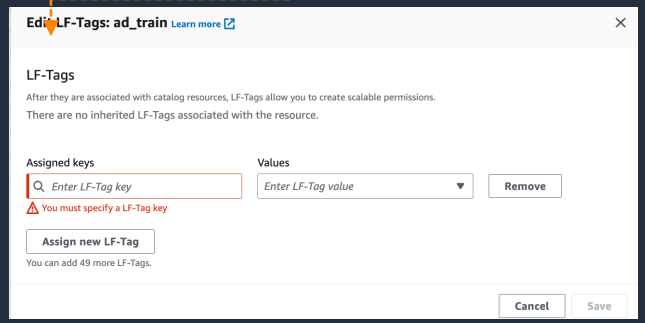




同時にそのタグを
誰が付与できるかも設定

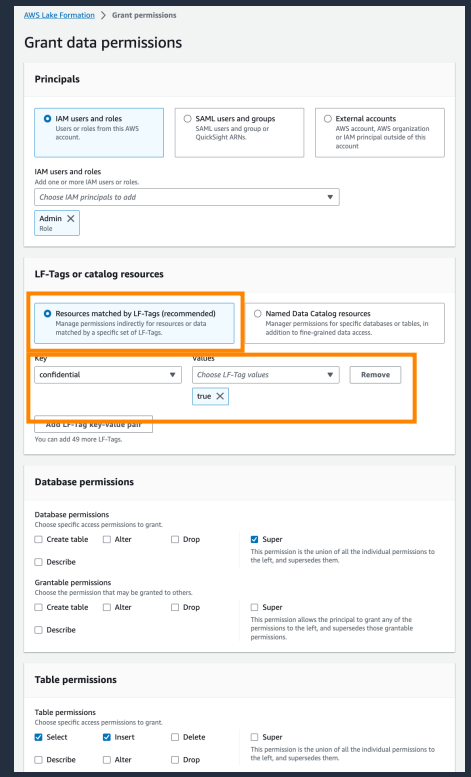
2. LF タグをリソース (DB/Table) に付与





DB/Table/Column に付与可能
タグは下位リソースに自動で継承
(DB のタグが Table/Column に継承)
継承されたタグは上書き可能

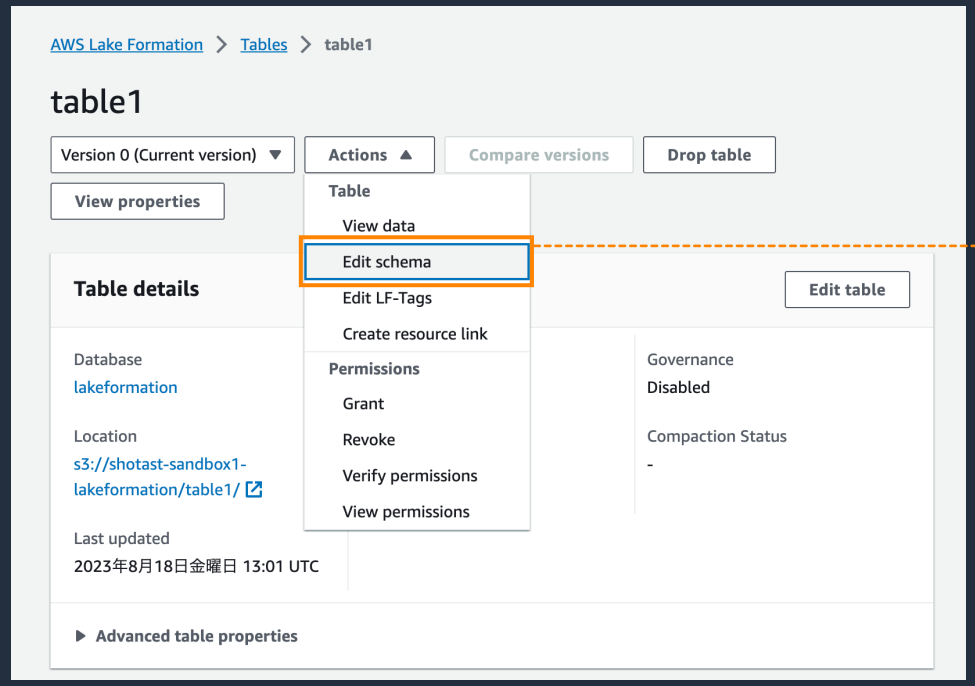
3. LF タグを プリンシパルに付与

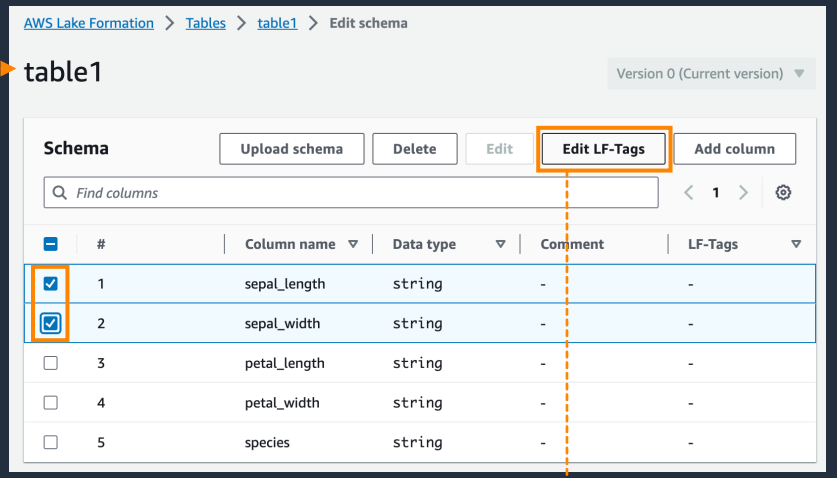


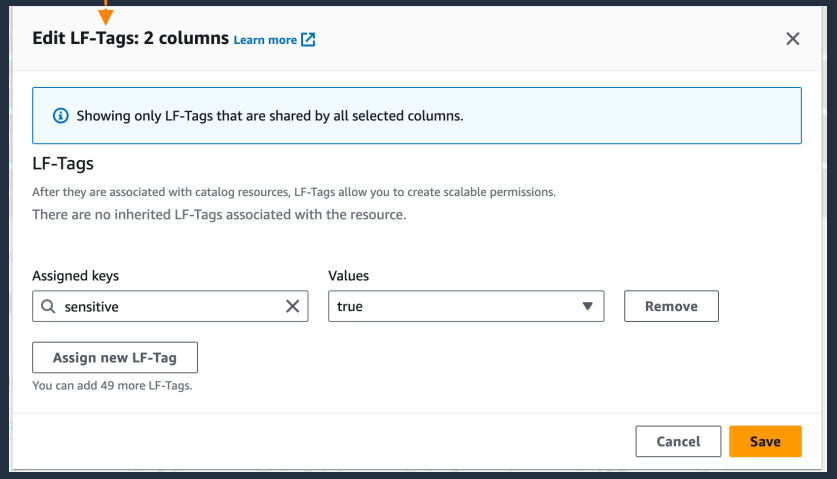
プリンシパルにタグと
権限を付与

LF タグによる列レベルセキュリティ

LF タグは Table の列にも付与ができるため、列に異なる LF タグを付与することで列レベルのセキュリティを実施できる

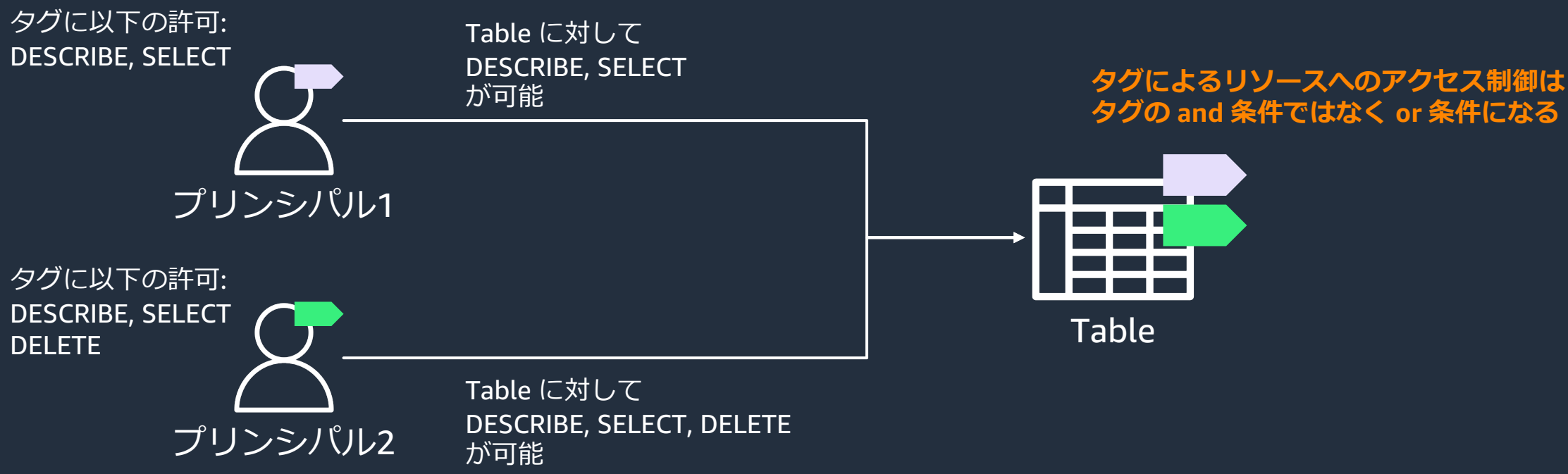






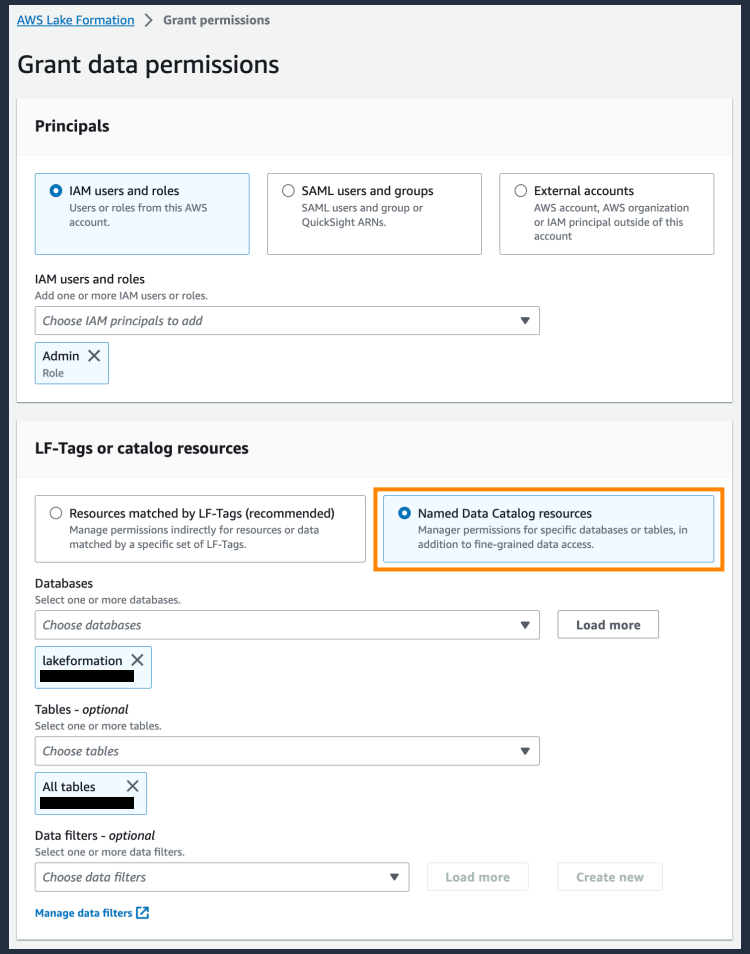
LF タグが複数付与されている場合の挙動

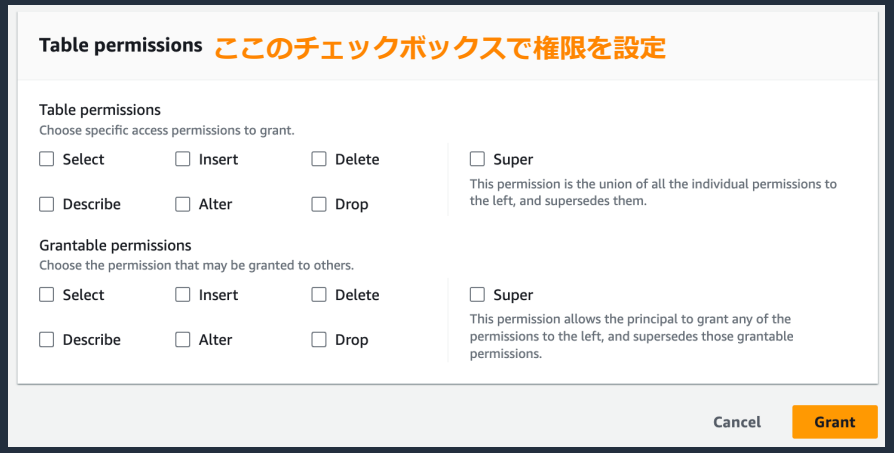
LF タグはリソースに複数付与することができるが、対応するタグがプリンシパルに1つでも付与されてれば、その権限でリソースにアクセスができる



リソース方式の設定方法

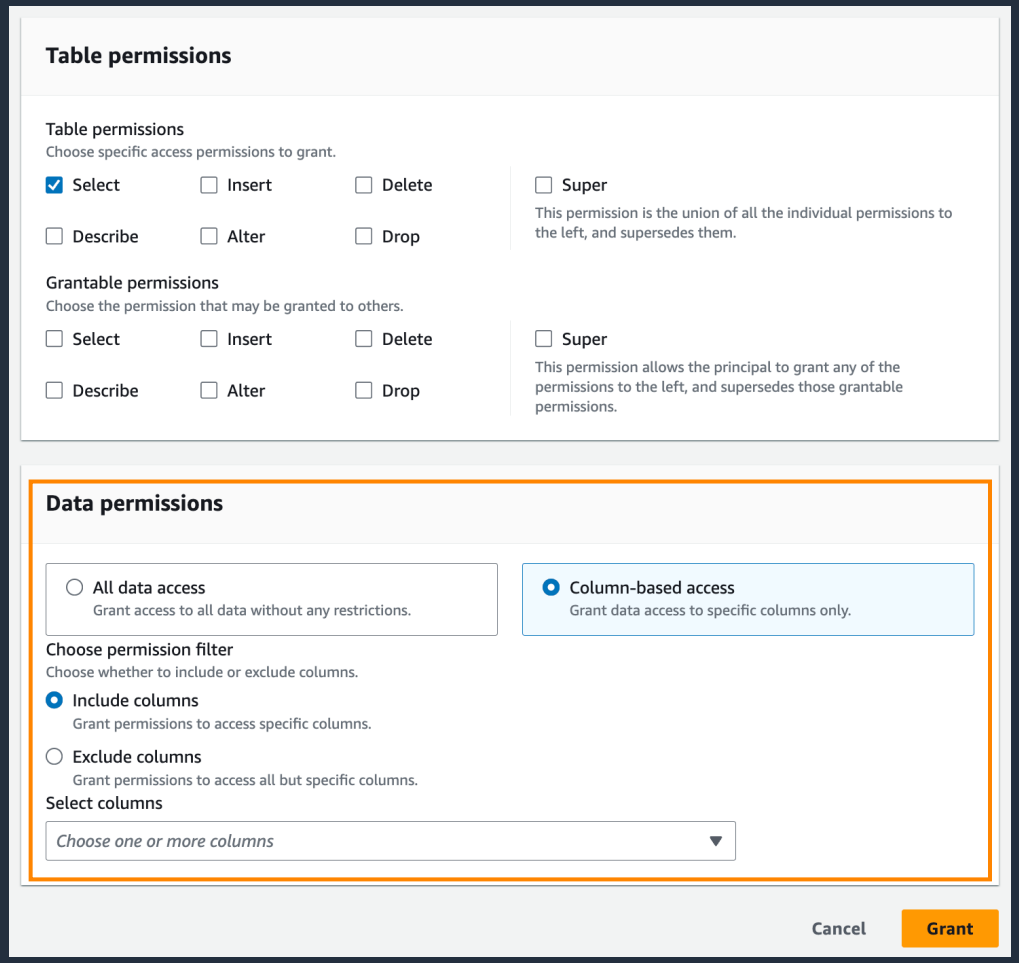
誰が (プリンシパル) どのリソース (DB/Table) にアクセスできるか直接設定する





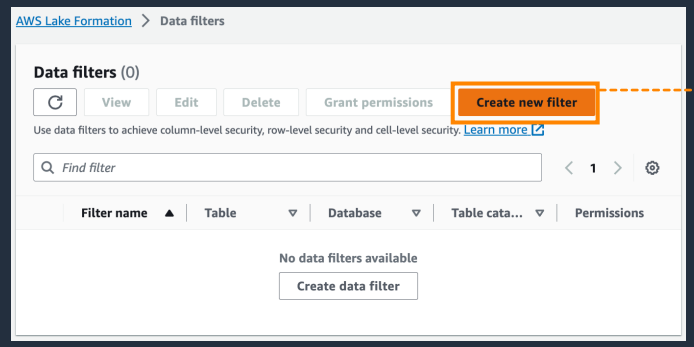
リソース方式のシンプルな列フィルタリング

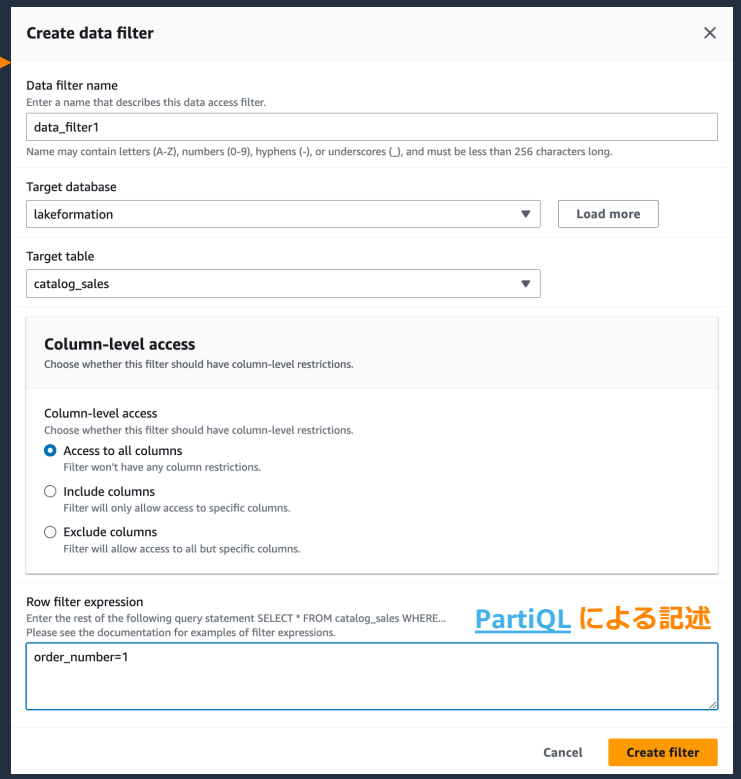
[Grant data permissions] でプリンシパルとテーブルの組み合わせに権限を付与する際に、アクセスできる列を指定できる

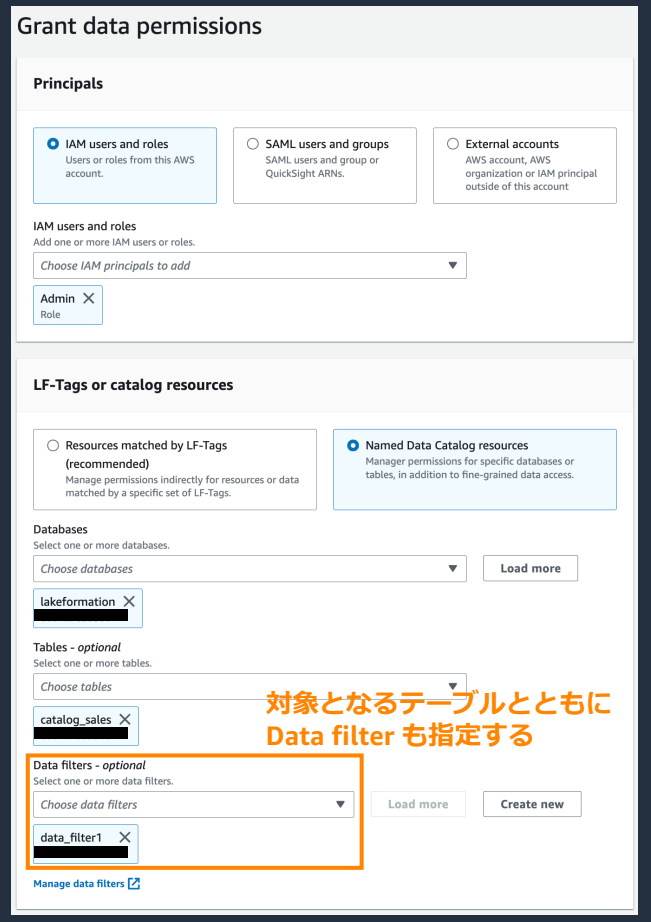


データフィルタによる行/列/セルレベルセキュリティ

リソース方式の設定ではデータフィルタを設定することができ、このデータフィルタでは行/列/セルレベルのセキュリティを設定できる

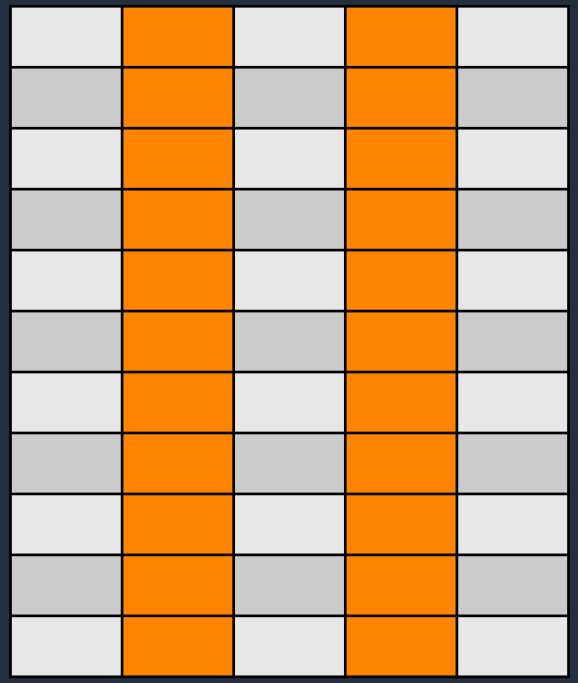






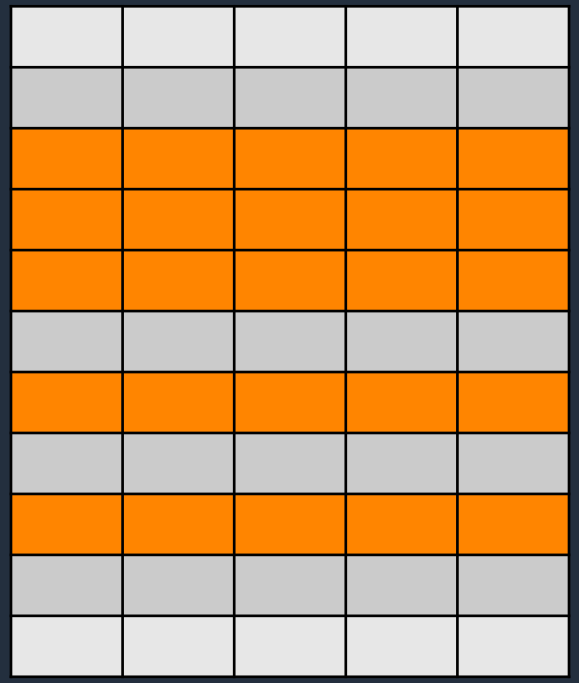
データフィルタによる行/列/セルレベルのセキュリティ イメージ

Columns



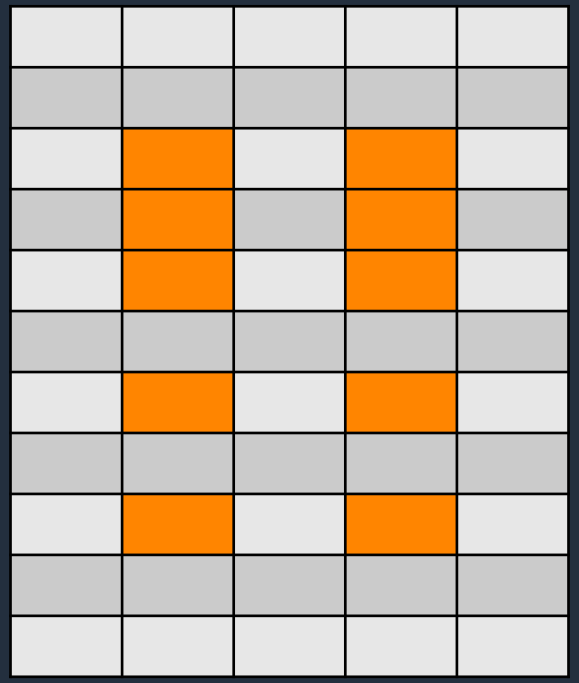
列を含む/除外する設定

Rows



PartiQL による行の設定

Cells



行と列の設定の組み合わせ

Amazon Athena, AWS Glue, Amazon Redshift Spectrum に対応

セキュリティレベルと AWS の各サービスの対応状況

サービスによって使えるセキュリティが異なる

メタデータ アクセス制御方式	制御方法	Amazon Athena	Amazon Redshift Spectrum	AWS Glue	Amazon EMR	Amazon QuickSight
LF タグ方式	LFタグによる 列フィルタリング	○	○	×	○	※2 ○
リソース方式	シンプルな 列フィルタリング	○	○	×	○	※2 ○
	データフィルタ	○	○	○	※1 △	※2 ○

※1 PartiQL による行レベルのセキュリティは無視される

※2 QuickSight は Athena を経由した場合のみ Lake Formation の制御を受ける

黙示的な Lake Formation の許可

明示的に権限を付与しなくても各ロールには権限が付与される

データレイク
管理者

- データカタログ内のすべての Describe リソースへのアクセス
- データレイク全体に対するデータロケーションへのアクセス許可
- データカタログ内の任意のリソースへのアクセス権を任意のプリンシパルに付与可能
- データベース作成権限を別プリンシパルに付与

データベース
作成者

- 作成するデータベースに対するデータベース許可とそこに作成されるテーブルに対する許可
- データベース内のテーブル作成権限を別プリンシパルに付与

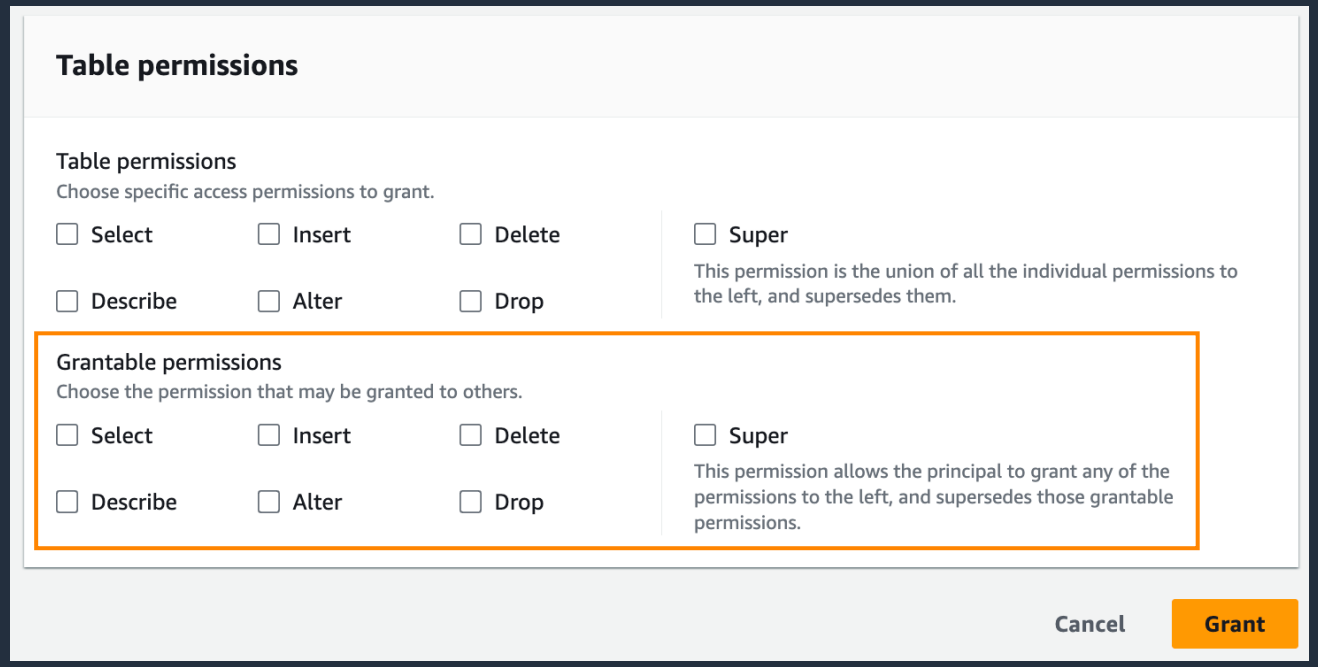
テーブル
作成者

- 作成するテーブルに対するすべての許可
- 作成するテーブルに対するアクセス権を別のプリンシパルに付与
- 作成するテーブルが含まれるデータベースの表示

https://docs.aws.amazon.com/lake-formation/latest/dg/implicit-permissions.html

権限付与権限の付与

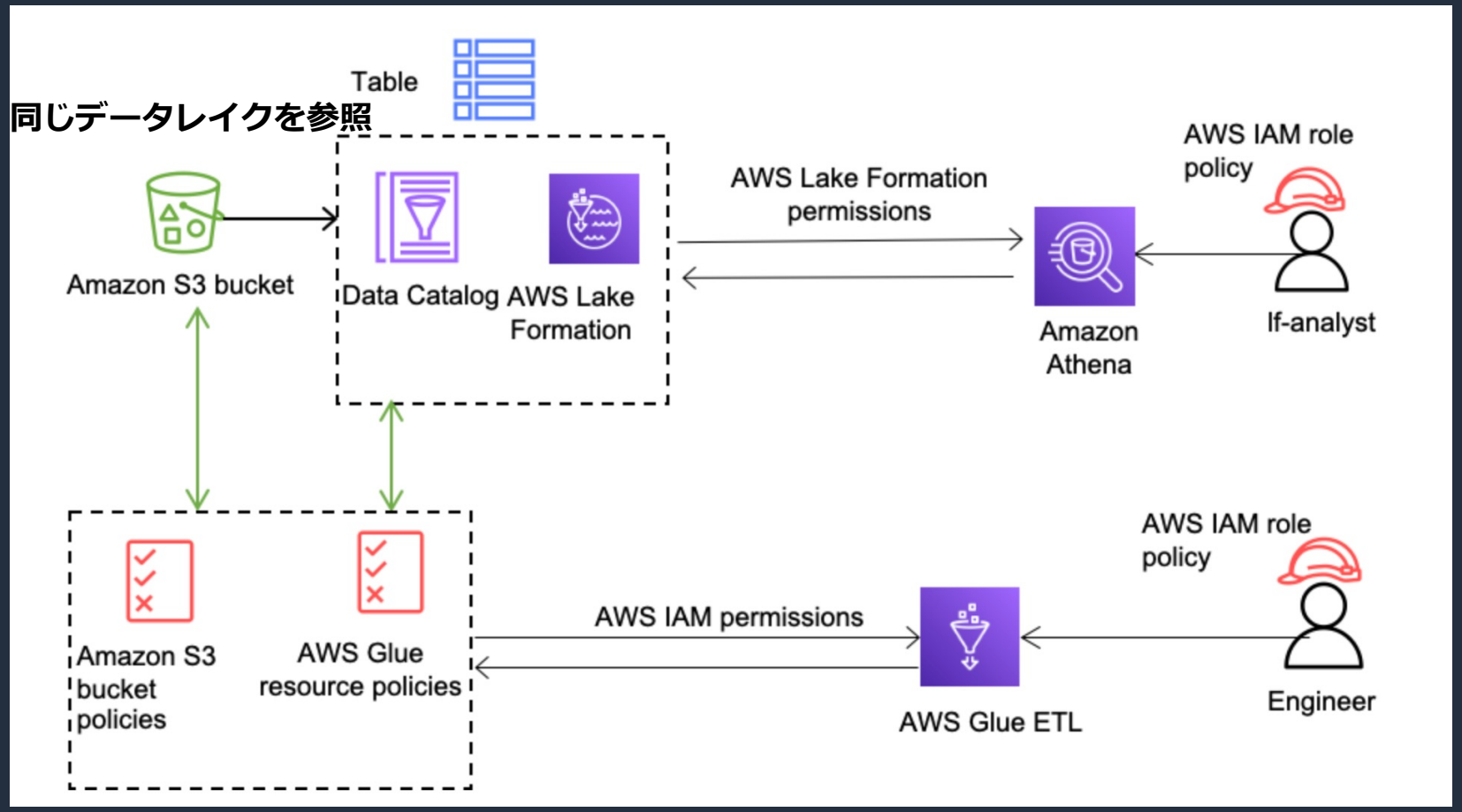
LF タグ方式/リソース方式ともに他のプリンシパルに権限を付与するための権限付与権限を付与することができ、それによってデータレイク管理者に依存せずにアクセス権限の制御が可能



対象のリソースに対して SELECT/INSERT /DELETE/DESCRIBE/ALTER/DROP の権限をプリンシパルに付与できる権限を付与する

Hybrid access mode

標準の IAM による制御と Lake Formation による制御を両立させることができるモード



lf-analyst は
LF の制御を受ける

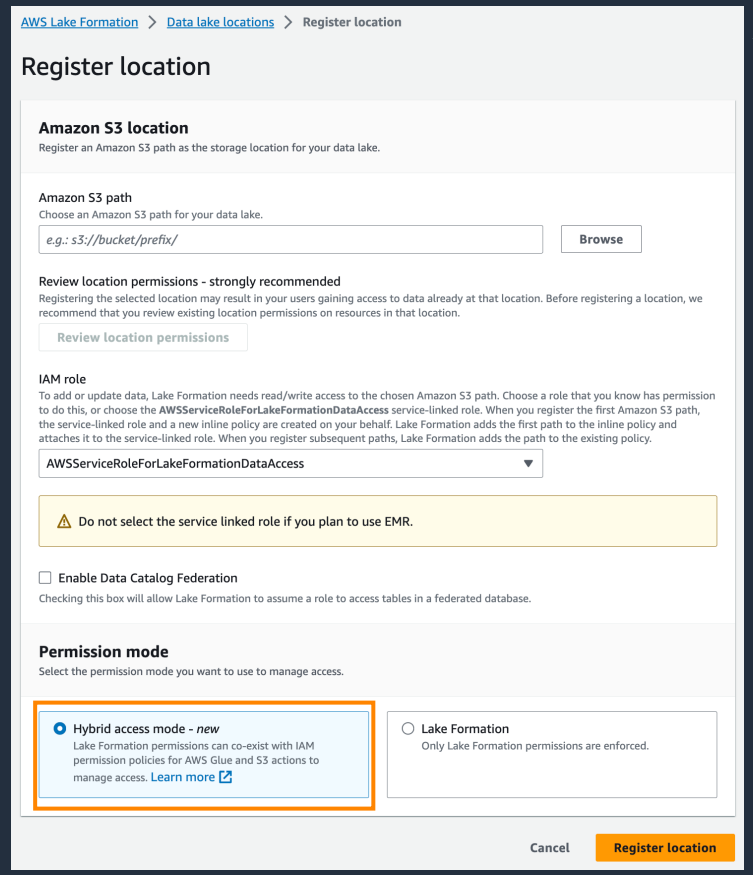
Engineer は
LF の制御を受けない

https://docs.aws.amazon.com/lake-formation/latest/dg/hybrid-access-mode.html

Hybrid access mode の設定方法

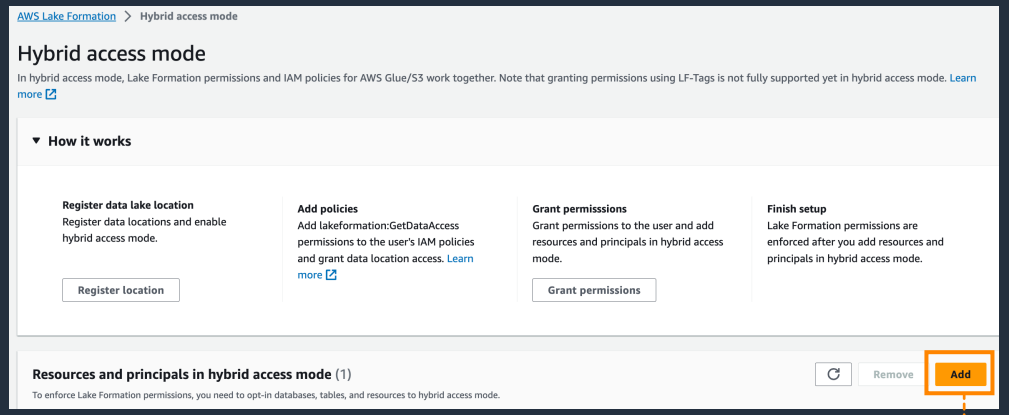
2つのステップで設定をおこなう

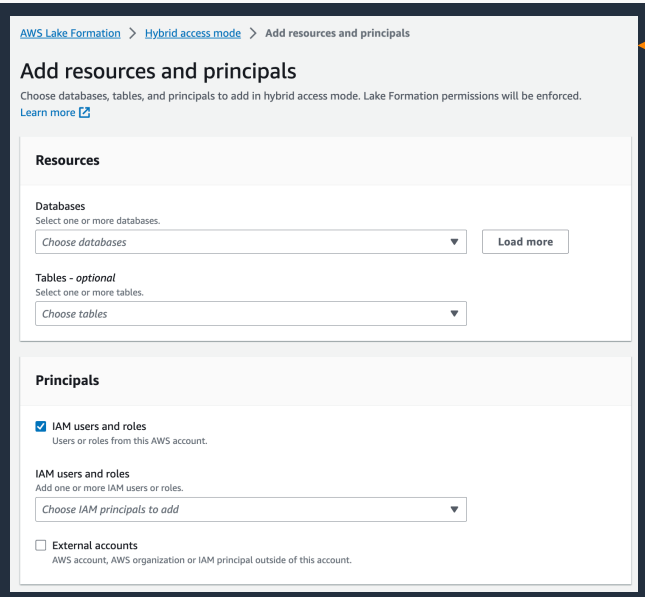
1. データレイクを Hybrid access mode で登録



登録済みのものは
[Data lake locations] → [Actions] → [Edit] で編集する

2. プリンシパルの制御を Lake Formation で行うように強制



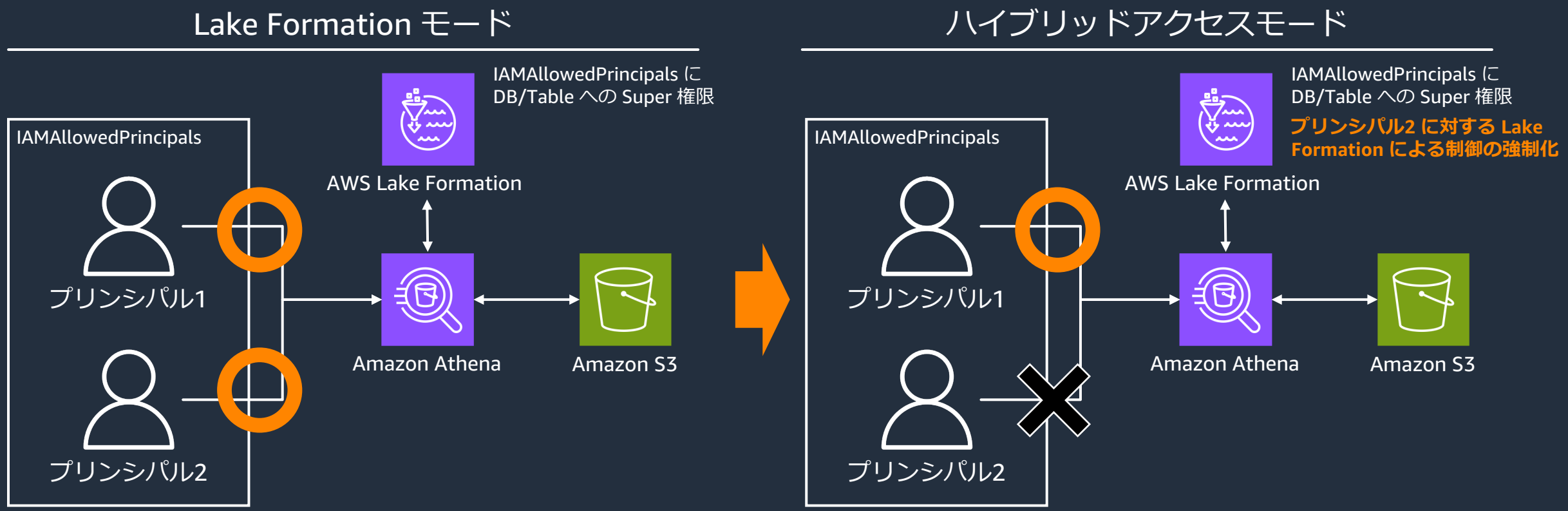


LF 制御を強制するリソースとプリンシパルの対象を選択する

LF タグによる制御を利用してる場合も同様に強制するリソースとプリンシパルを選択する

Hybrid access mode の仕組み

Hybrid access mode は IAMAllowdPrincipals が付与されていても Lake Formation による制御を強制的に行わせる機能

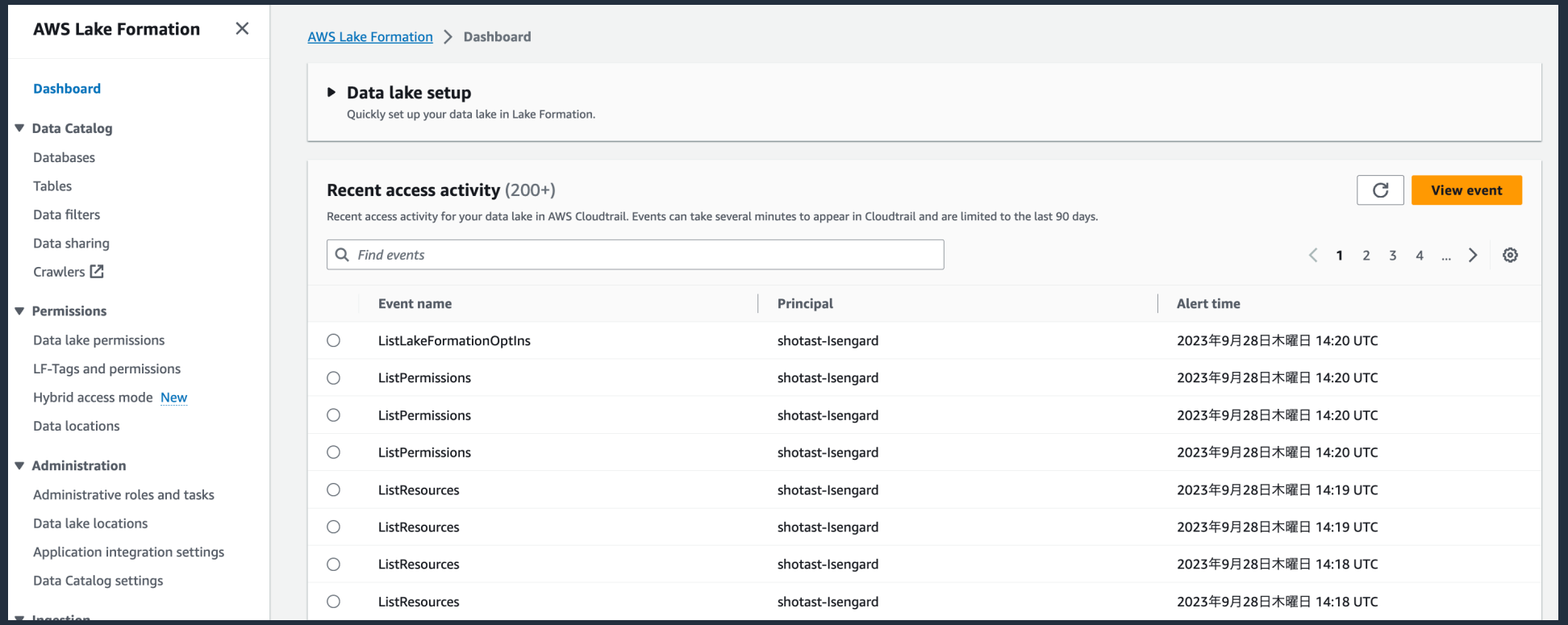


IAMAllowedPrincipals に Super 権限が付与されてるため
IAMAllowedPrincipals に属するプリンシパルは
任意の DB/Table にアクセス可能

Lake Formation による制御が強制化されてるため
プリンシパル2 は IAMAllowedPrincipals に属しながら
それとは別に Lake Formation 上で
メタデータのアクセス許可が必要

CloudTrail との統合

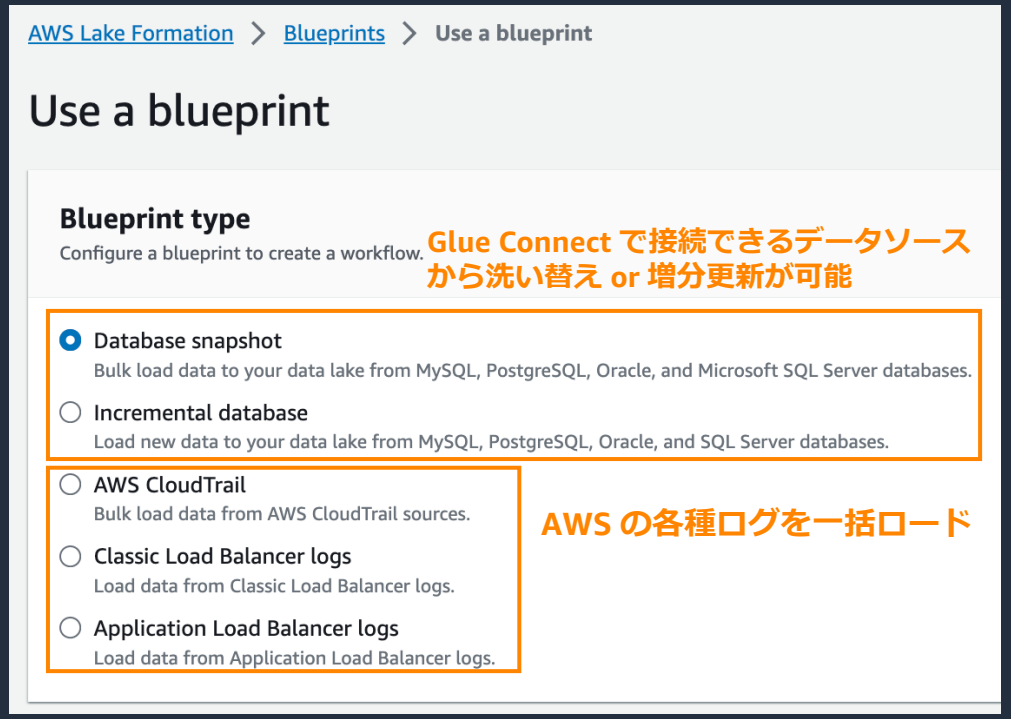
Lake Formation は CloudTrail と統合されており、Lake Formation API のコールをイベントとしてキャプチャする。Lake Formation のコンソール画面で確認が可能

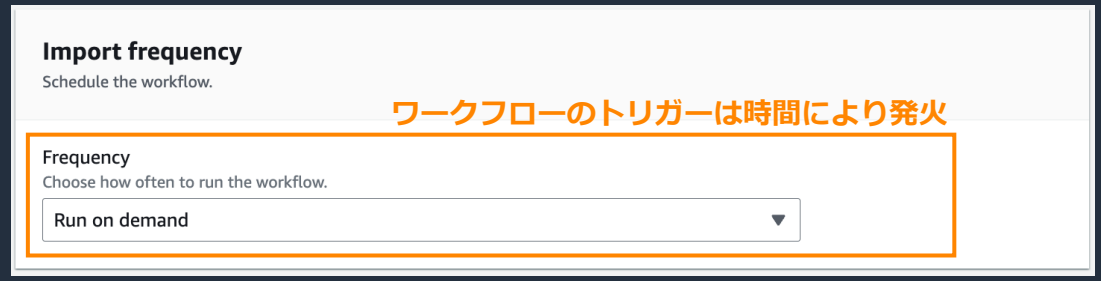


AWS Lake Formation の機能 - ブループリントの概要と機能

ブループリント機能とは？

ブループリントはデータレイクにデータを取り込むためのワークフローを手軽に実装できる機能で、様々な種類がある

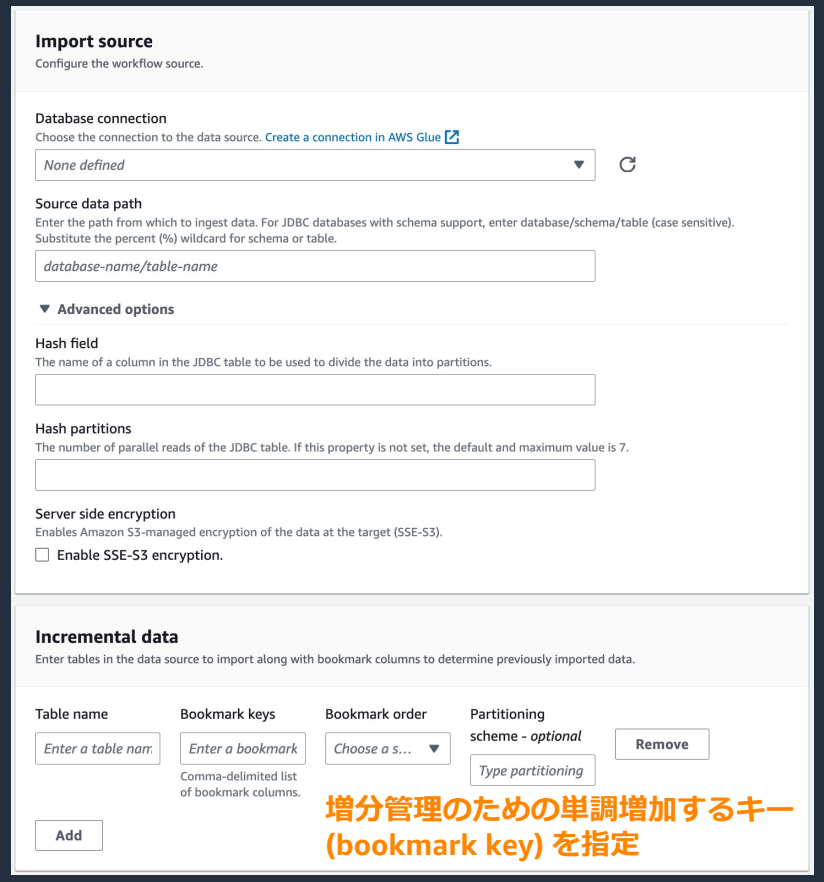




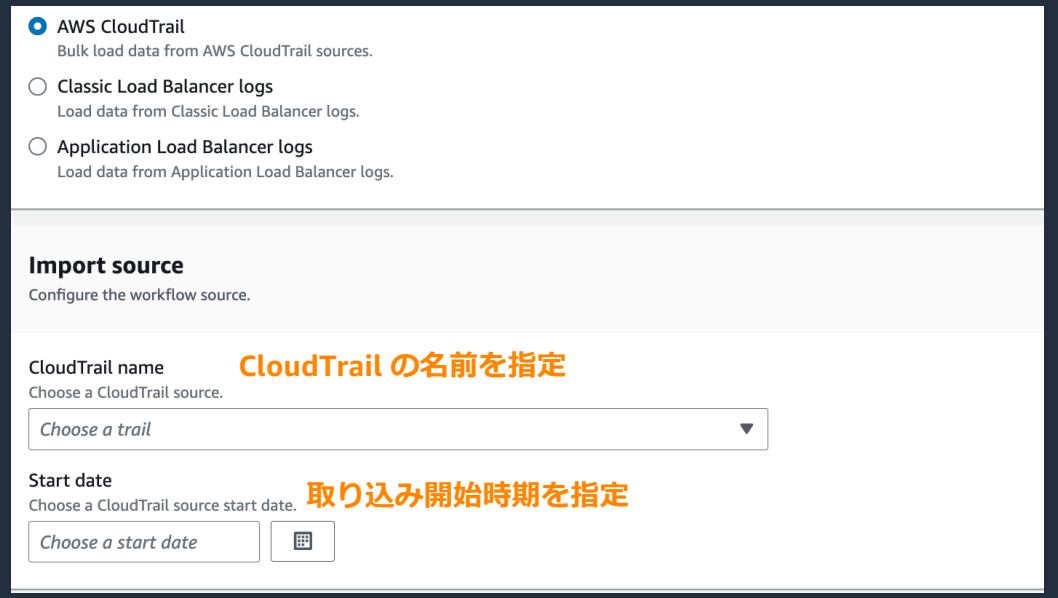
ブループリントの作成例

それぞれの種類によって必要な設定が異なる

Incremental Database (増分更新)



AWS CloudTrail



補足：ブループリント vs AWS Glue

ブループリントは AWS Glue workflow でトリガ、ジョブ、クローラーを自動的に設定するため、実装の難易度が低いのがポイント。AWS Glue はブループリントよりも遥かに実装の自由度が高い。Glue で実装ができるのであれば、障害対応や実装の透明性の観点で Glue を推奨

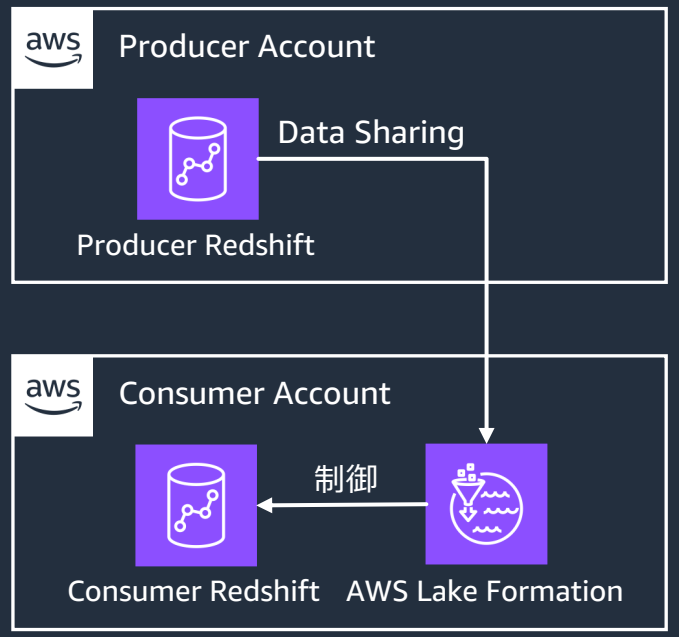
	ブループリント	AWS Glue
GUI での実装	可能 (CLI にはなし)	AWS Glue Studio を使えば可能
Spark コードの実装有無	不要 （ブループリントが全て用意）	場合によっては必要 （Glue Studio を使うと自動で生成できる）
ワークフローの実装有無	不要 （ブループリントが全て用意）	必要で、AWS Glue workflow や AWS Step Functions などを利用
トリガーの自由度	時間や頻度のみ （裏でAWS Glue workflowが作られるので それを編集すれば他の Glue Job や EventBridge と連携可能）	使うワークフローエンジンに依存する

AWS Lake Formation の機能 - データ共有の概要と機能

AWS Lake Formation によるデータ共有の種類

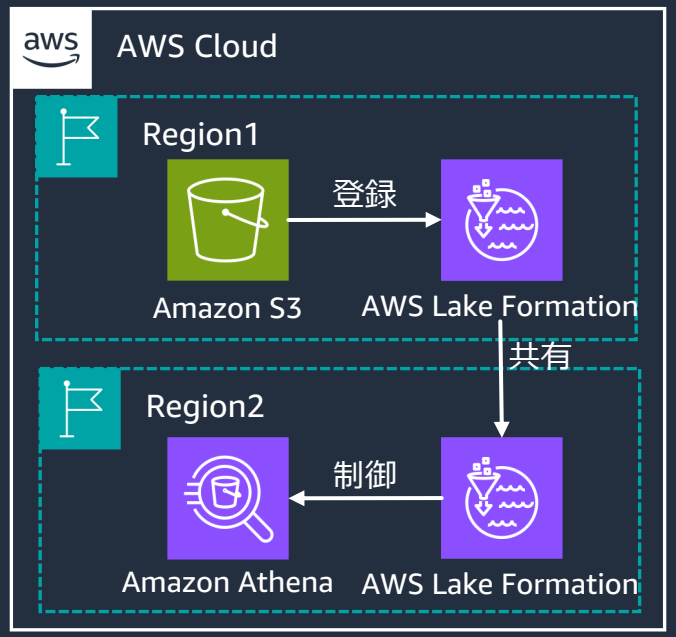
Lake Formation は3種類のデータ共有を制御することが可能

Redshift Data Sharing



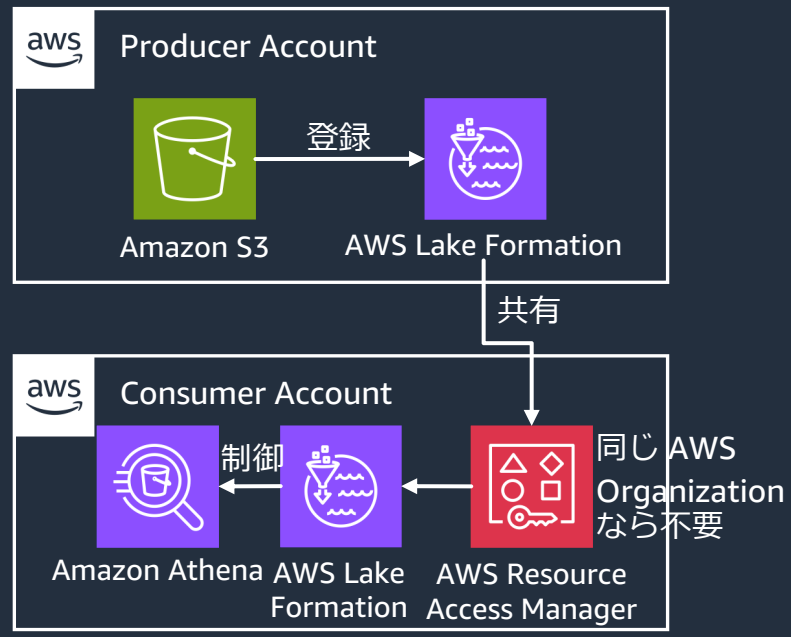
共有された Redshift のデータを
Consumer 側で制御

クロスリージョン共有



同じアカウントだが
リージョンを跨いだ
データレイクの共有と制御が可能

クロスアカウント共有

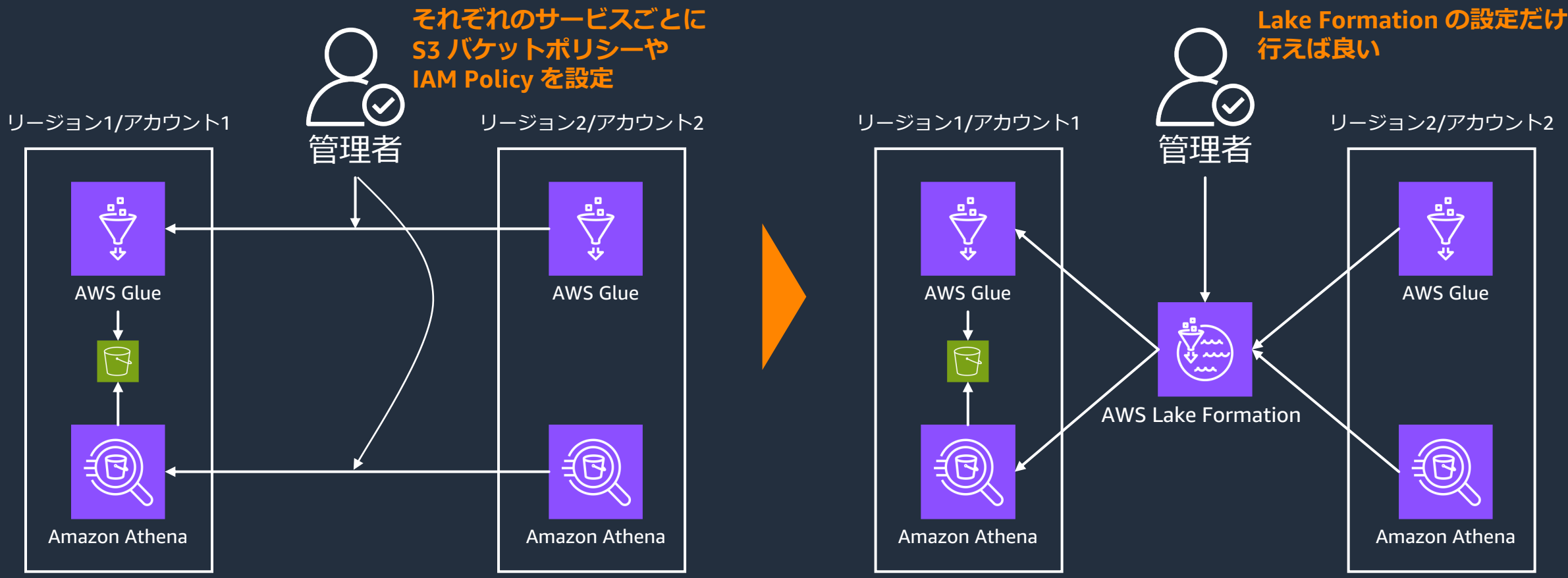


同じリージョンだが
アカウントを跨いだ
データレイクの共有と制御が可能

組み合わせて利用することでリージョンと
アカウントを跨いでデータレイクの共有が可能

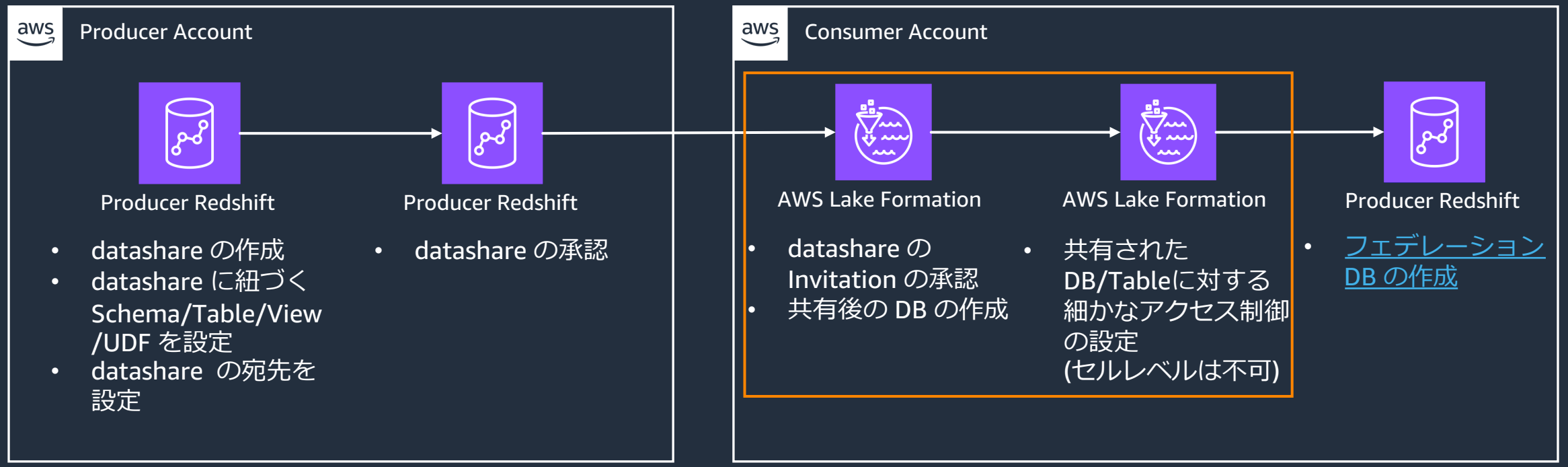
AWS Lake Formation を利用する理由

クロスリージョン/クロスアカウント共有は Lake Formation を使わなくても実現可能だが、
Lake Formation を使うことで一元的に設定が可能なため、運用負荷が減る



Redshift Data Sharing における Lake Formation の制御

Redshift Data Sharing を行う際に Consumer 側で Lake Formation による制御を行い、共有された側で細かな粒度のアクセス制御を実現できる



https://docs.aws.amazon.com/lake-formation/latest/dg/data-sharing-redshift.html

クロスリージョン共有

リソースリンクを作ることで、同一アカウント他リージョンへのリソースを共有することが可能

1. S3 のリソースを共有元リージョン (Producer) で Lake Formation に登録して細かな粒度のアクセス制御を実施

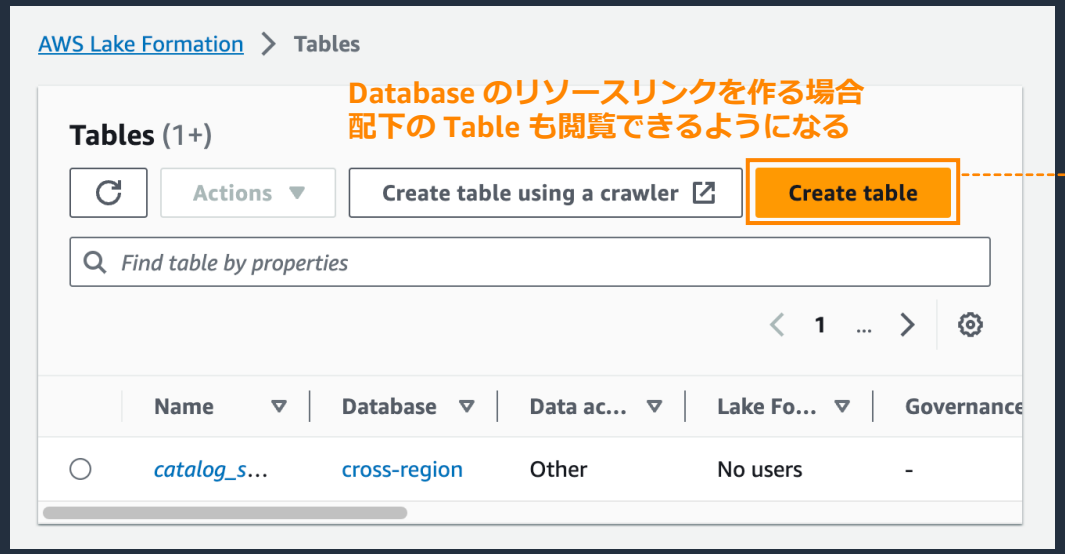
2. 共有先リージョン (Consumer) で Producer の資源を指定して Lake Formation 上でリソースリンクを作成

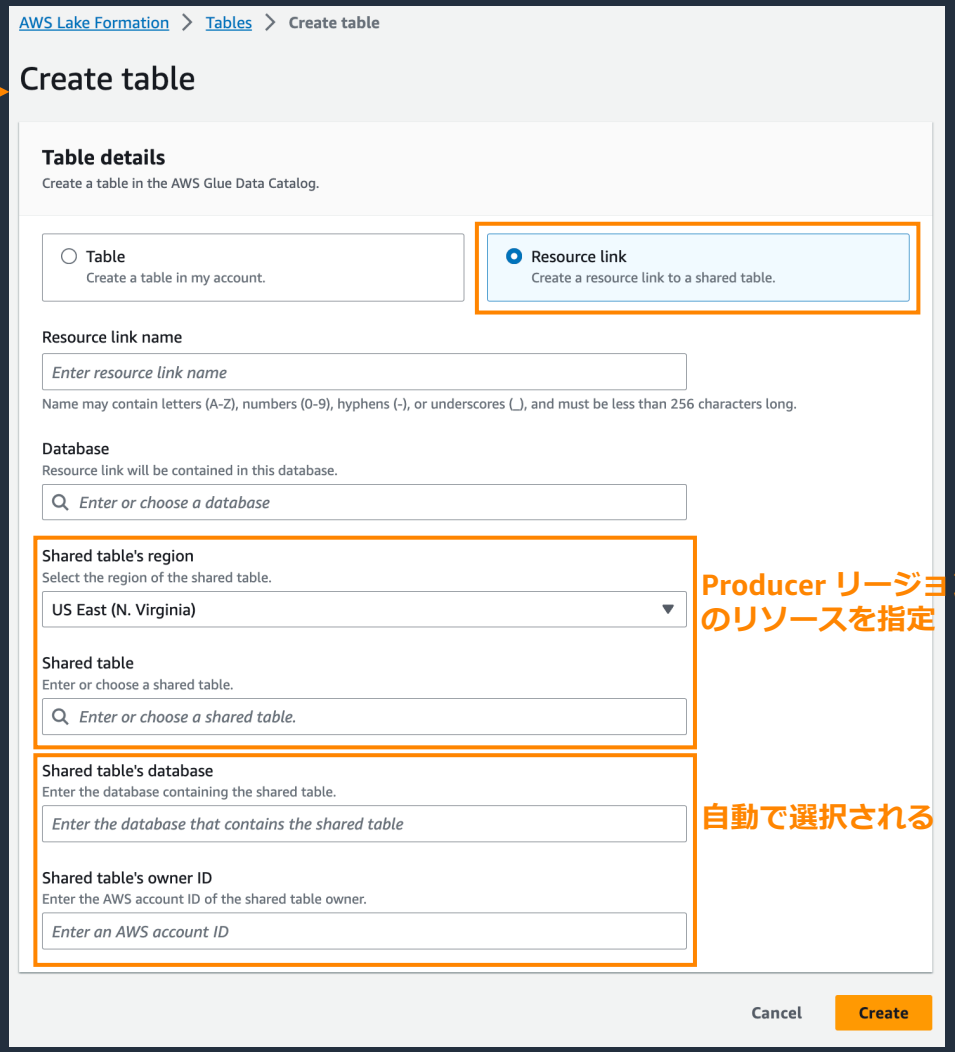
3. Consumer 側で Lake Formation のアクセス制御を利用して Athena 等のサービスでアクセス制御を実施

※ Consumer 側では [Describe]/[Drop] の権限しか設定できないため、
見せる見せないの制御しかできない。細かな粒度の制御は Producer 側で実施する

クロスリージョン共有 – リソースリンクの作成

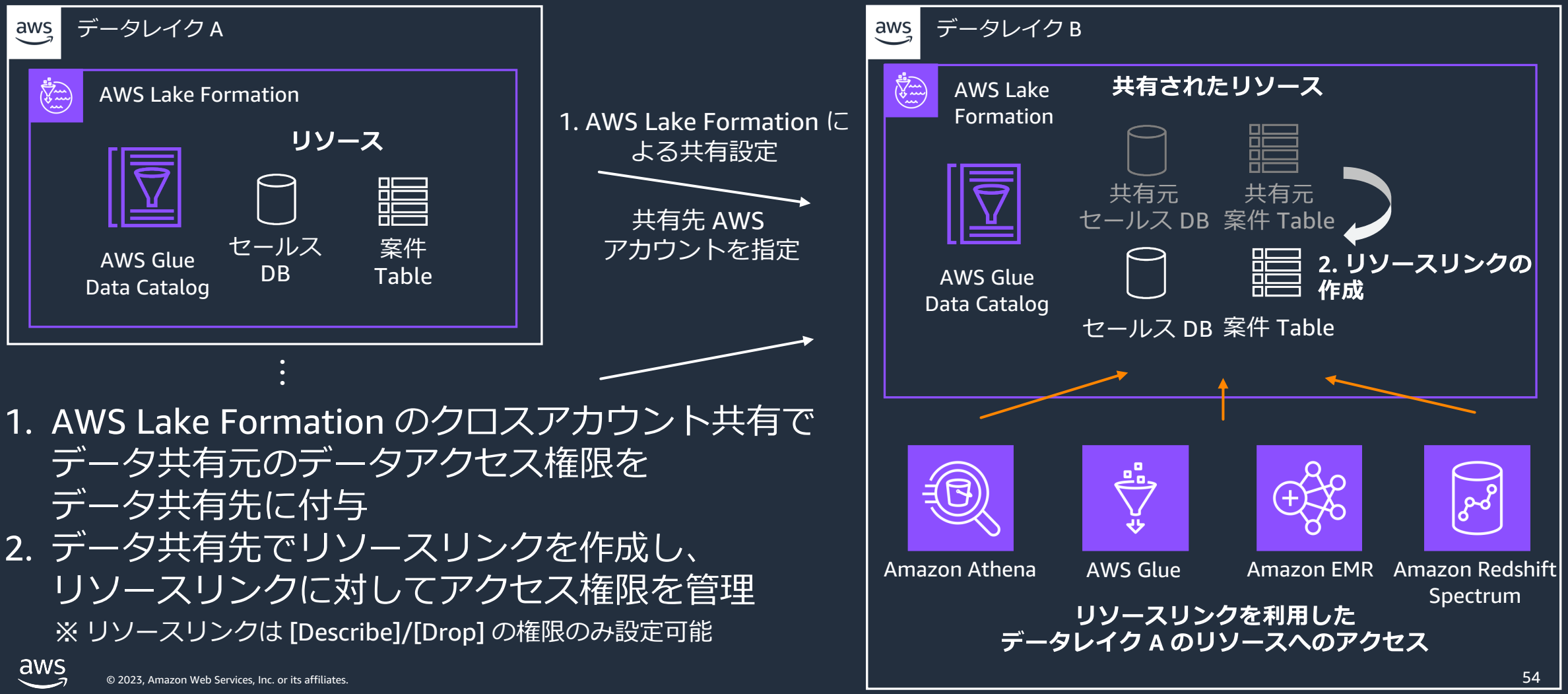
共有先 (Consumer)のリージョンで [Databases]/[Tables] → [Create database]/[Create table] から作成が可能





クロスアカウント共有

Lake Formation による細かな粒度のアクセス制御を実施した状態で、他 AWS アカウントにデータレイクを共有可能（LFタグ方式、リソース方式どちらも可能）



1. AWS Lake Formation のクロスアカウント共有でデータ共有元のデータアクセス権限をデータ共有先に付与
2. データ共有先でリソースリンクを作成し、リソースリンクに対してアクセス権限を管理

※ リソースリンクは [Describe]/[Drop] の権限のみ設定可能

クロスアカウント共有 – Producer 側の設定（前提条件）

クロスアカウント共有を始めるためには権限の整理や AWS Glue のリソースポリシーの変更が必要

1. 共有するリソースに対して IAMAllowedPrincipals 許可をすべて排除する
   - IAMAllowedPrincipals があると AccessDeniedException が発生する
2. AWS Glue リソースポリシーの更新
   - 以下の JSON を Glue API で実行

policy.json

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

リソース方式に必要な設定
<region> と <source-account-id> に適切な値を入れる

LF タグ方式に必要な設定
<target-account-id>, <region> と <source-account-id> に適切な値を入れる

```
aws glue put-resource-policy --policy-in-json
file://policy.json --enable-hybrid TRUE
```

※ ファイルを指定する場合 file:// が必要

クロスアカウント共有 – Producer の設定①

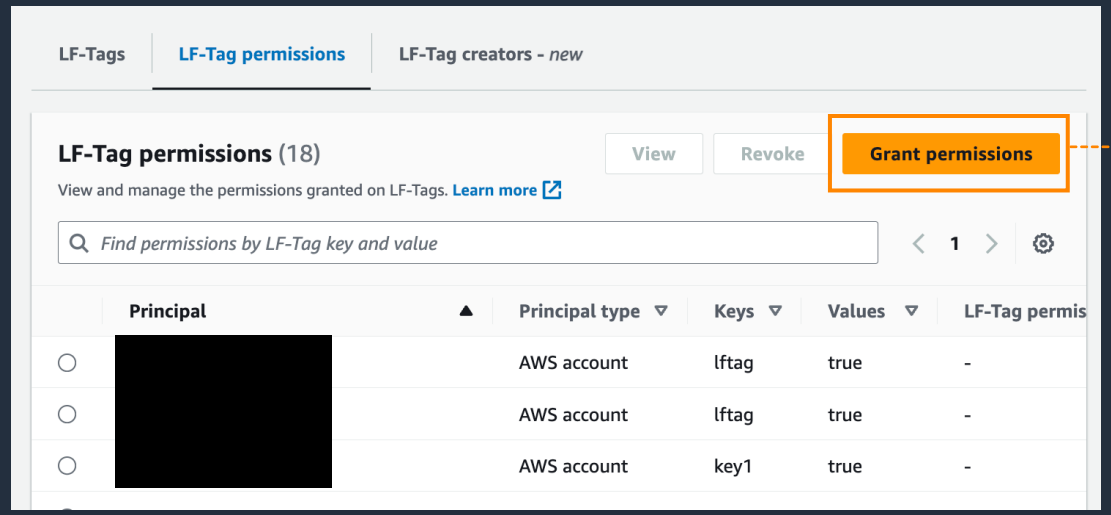
[Data Catalog settings] → [Cross account version settings] で、クロスアカウントの仕様バージョンを切り替えられるが、基本最新の Version4 を使えば良い

Producer と Consumer は同じバージョンにしておくのが良い
Producer が v1, v2 で Consumer が v3, v4 だとエラーが発生する
(Invalid cross account grant request)

クロスアカウント共有 – Producer の設定②

LF タグ方式のアクセス制御で共有を行う場合は、Consumer 側に LF タグの権限を付与する（リソース方式では不要）

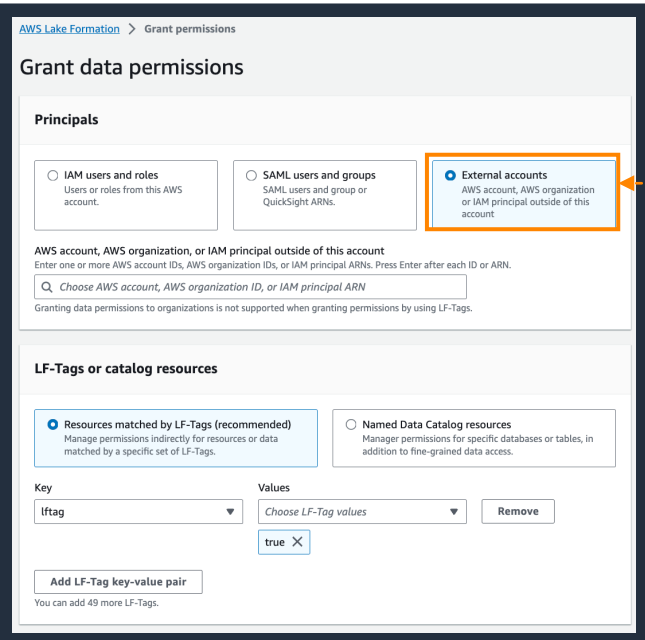
[LF-Tags and Permissions]



クロスアカウント共有 – Producer の設定③

[Data lake permissions] で Consumer の AWS Account/IAM Role を指定して、方式に応じてアクセス制御を行い、共有

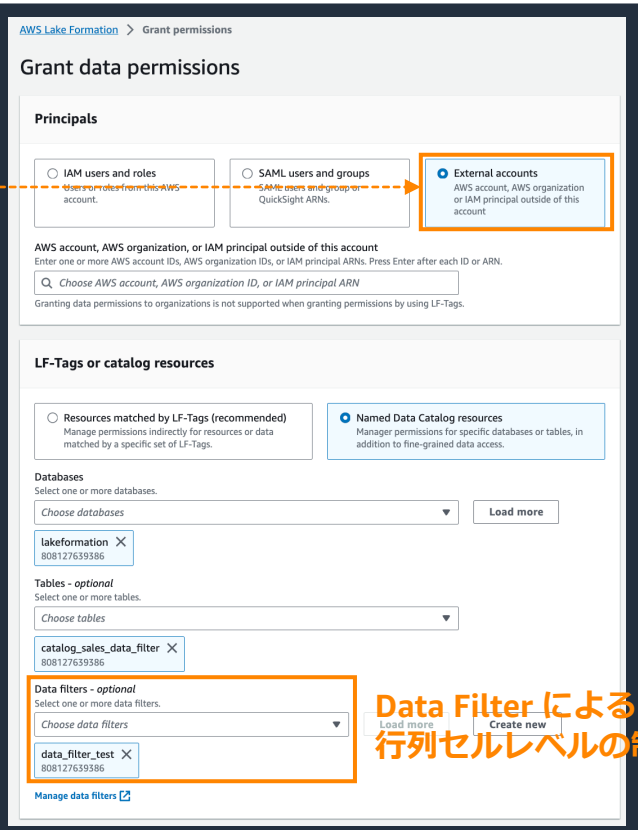
LF タグ方式



LF タグ方式の共有では LF タグによる
列レベルセキュリティを実施可能

どちらも External accounts を設定

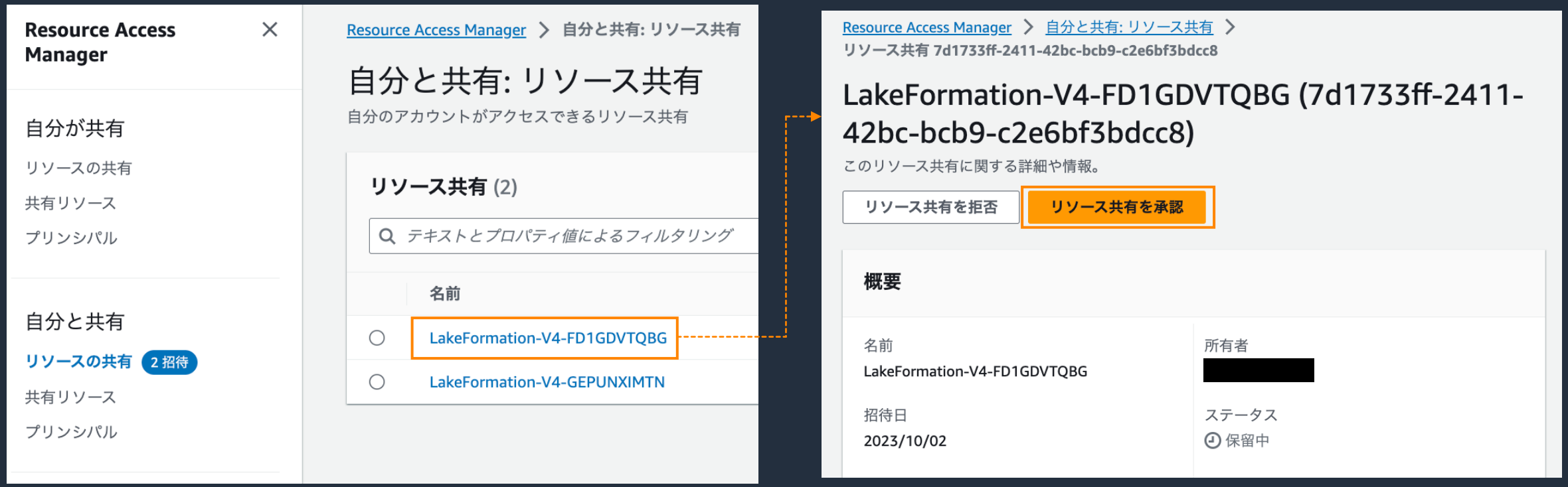
リソース方式



Data Filter による
行列セルレベルの制御が可能

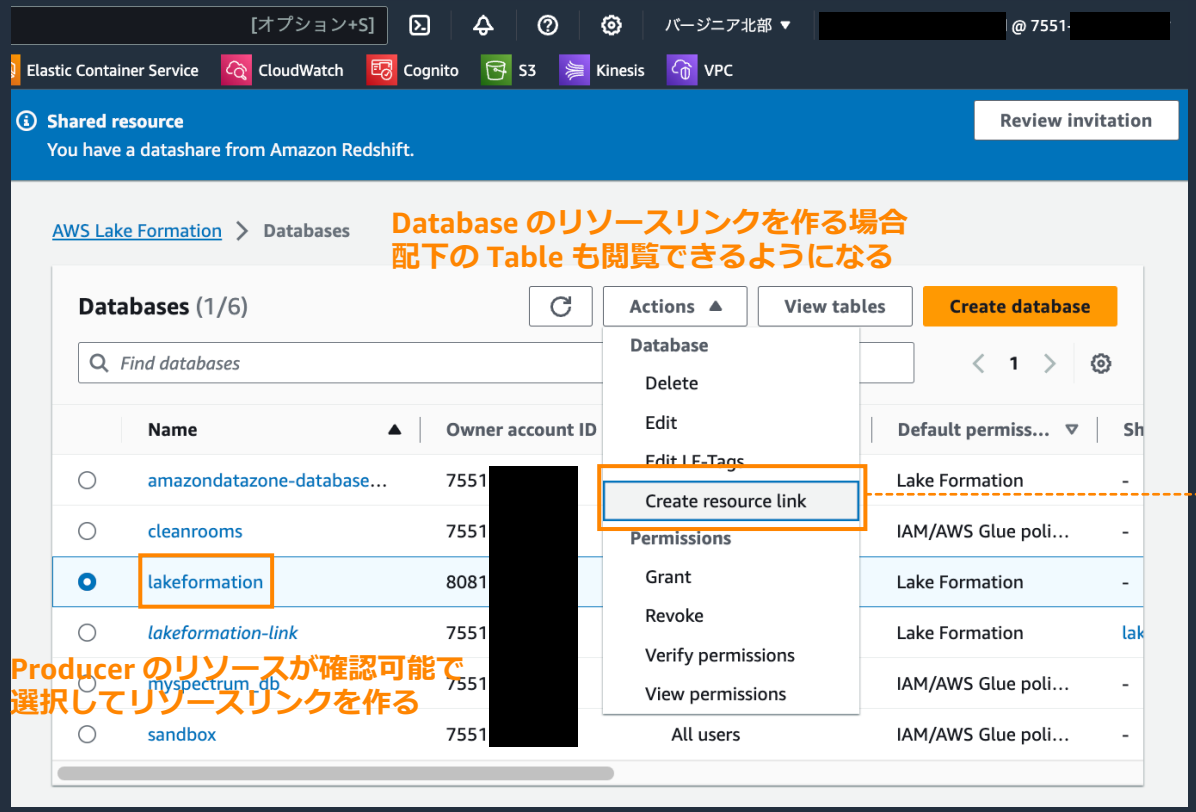
クロスアカウント共有 – Consumer 側の設定①

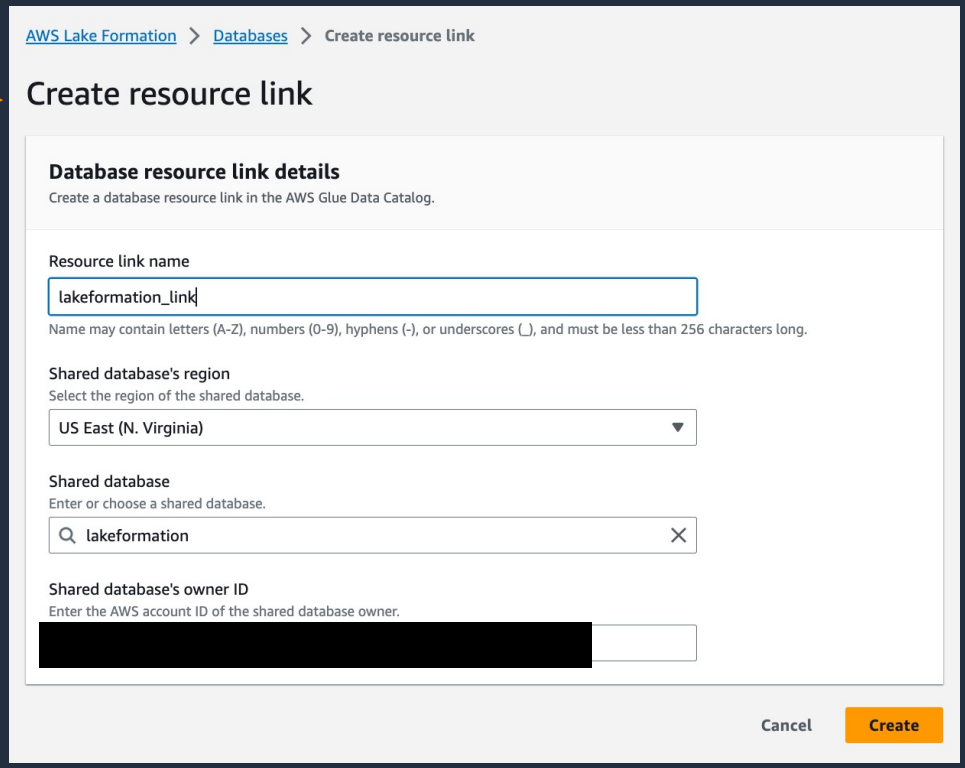
同じ AWS Organization に属してない場合、AWS RAM での受け入れ承認が必要



クロスアカウント共有 – Consumer 側の設定②

共有されたリソースは Lake Formation のコンソールからは確認できるが、Athena 等から利用できないため、リソースリンクを作成する





AWS Lake Formation の料金

AWS Lake Formation 費用

AWS Lake Formation 自体の利用は無償

配下で活用するサービスに

かかる費用のみのお支払い

まとめ

まとめ

- 細かな粒度のアクセス制御を様々な AWS サービスに実施可能
  - アクセス制御の裏側、制御の対象、IAMAllowedPrincipals などLake Formation の権限管理の仕組みを紹介
  - LF タグによるアクセス制御とリソース方式によるアクセス制御を説明
- ブループリントによる簡単なデータ取り込み
  - DB, AWS ログの自動取り込み、Glue のトリガー、ジョブ、クローラーの自動設定
- Lake Formation を利用した安全なデータ共有
  - Redshift Data Sharing の Consumer 側での制御、クロスリージョン共有、クロスアカウント共有が可能
- Lake Formation 自体の費用は無償
  - 配下で仕様するサービス費用のみ

参考資料

- AWS Lake Formation 公式ドキュメント
  - https://docs.aws.amazon.com/lake-formation/latest/dg/what-is-lake-formation.html
- Amazon Redshift “What is datashare?”
  - https://docs.aws.amazon.com/redshift/latest/dg/what_is_datashare.html#lf_datashare_overview
- AWS 公式ブログ “Lake Formation を使用し AWS アカウント間でセキュアにデータ共有を実現”
  - https://aws.amazon.com/jp/blogs/news/lake-formation-cross-account/
- AWS CLI Command Reference “glue”
  - https://docs.aws.amazon.com/cli/latest/reference/glue/#cli-aws-glue

Thank you!