



AWS
Black Belt
Online Seminar

このコンテンツは公開から3年以上経過しており内容が古い可能性があります
最新情報については[サービス別資料](#)もしくはサービスのドキュメントをご確認ください

【AWS Black Belt Online Seminar】

AWS Lake Formation

サービスカットシリーズ

アマゾン ウェブ サービス ジャパン株式会社

ソリューションアーキテクト 上原 誠

2019/10/01

自己紹介

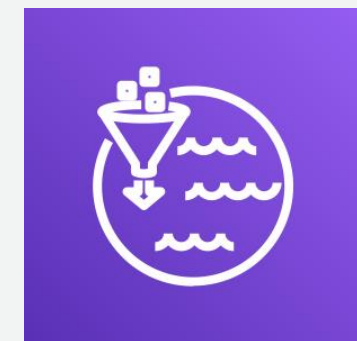
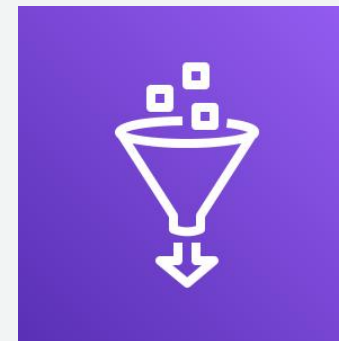
名前: 上原 誠 (うえはら まこと)

所属: ソリューションアーキテクト



好きな AWS のサービス : AWS Glue, AWS Lake Formation

担当 : メディア, アドテック etc



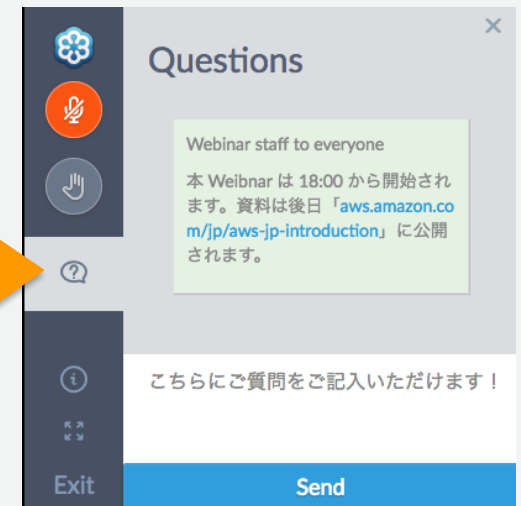
AWS Black Belt Online Seminar とは


「サービス別」「ソリューション別」「業種別」のそれぞれのテーマに分かれて、アマゾン ウェブ サービス ジャパン株式会社が主催するオンラインセミナーシリーズです。

質問を投げることができます！

- 書き込んだ質問は、主催者にしか見えません
- 今後のロードマップに関するご質問はお答えできませんのでご了承下さい

- ① 吹き出しをクリック
- ② 質問を入力
- ③ Sendをクリック



 Twitter ハッシュタグは以下をご利用ください
#awsblackbelt

内容についての注意点

- 本資料では2019年10月01日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(<http://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おります。日本居住者のお客様が東京リージョンを使用する場合、別途消費税をご請求させていただきます。
- AWS does not offer binding price quotes. AWS pricing is publicly available and is subject to change in accordance with the AWS Customer Agreement available at <http://aws.amazon.com/agreement/>. Any pricing information included in this document is provided only as an estimate of usage charges for AWS services based on certain information that you have provided. Monthly charges will be based on your actual use of AWS services, and may vary from the estimates provided.

Agenda

- データレイクとは
- AWS Lake Formation 登場の背景
- AWS Lake Formation とは
- AWS Lake Formation の機能
 - セキュリティ
 - データカタログ
 - ブループリント
- 落ち穂拾い
- 料金
- まとめ

データレイクとは



我々が考える以上に データは増えている

データ	データプラットフォームに 求められるもの	
成長率 > 10倍 5年毎	使用期間 15 年間	拡張性 1,000 倍

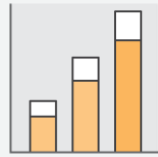
* IDC, Data Age 20215: The Evolution of Data to Life-Critical Don't Focus on Big Data, Focus on the Data That's Big, April 2017.



データサイエンティスト



ビジネスユーザー



分析



アプリケーション

セキュア リアルタイム

柔軟性

拡張性

データにアクセスする人々も増えている

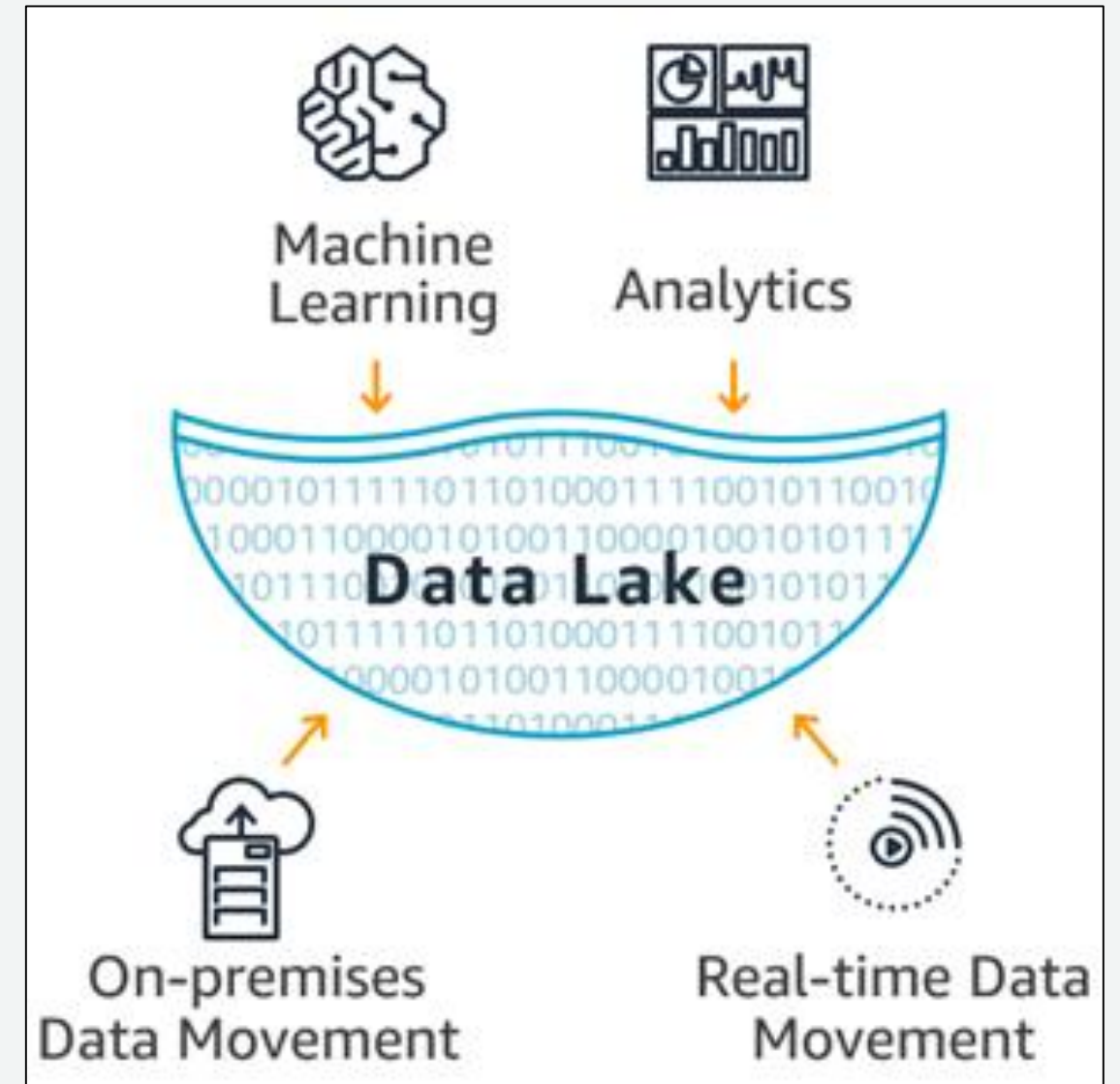
分析対象のデータに対する要件も増えている

データレイクとは

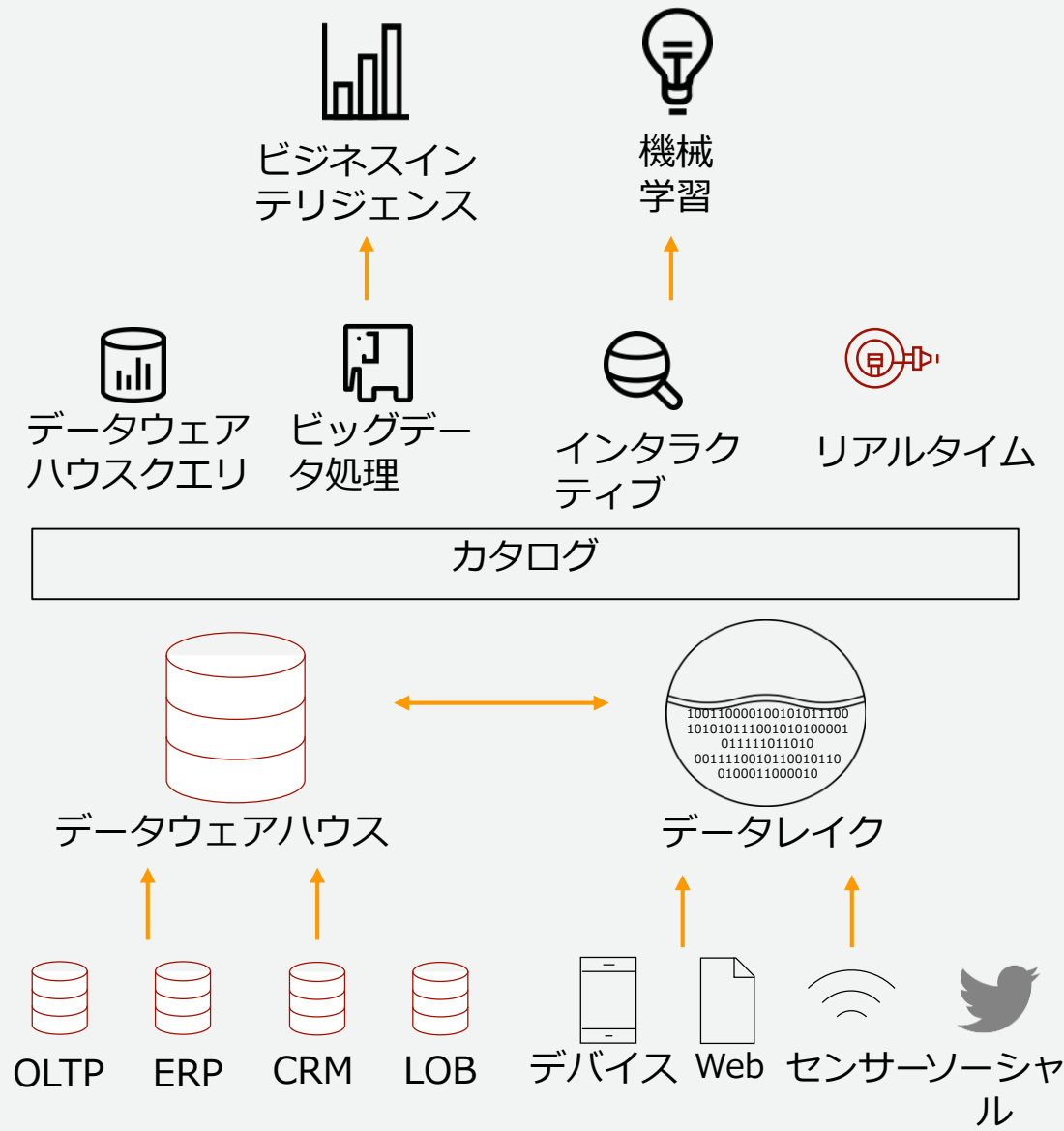
AWSの公式ドキュメント

<https://aws.amazon.com/jp/big-data/datalakes-and-analytics/what-is-a-data-lake/>

データレイクは、規模にかかわらず、**すべての構造化データと非構造化データを保存**できる一元化されたリポジトリです。データをそのままの形で保存できるため、データを構造化しておく必要がありません。また、ダッシュボードや可視化、ビッグデータ処理、リアルタイム分析、機械学習など、**さまざまなタイプの分析を実行**し、**的確な意思決定に役立てる**ことができます。



なぜデータレイクが必要か



データレイクが提供するもの:

構造化、半構造化、非構造化データの取り扱い

ペタバイト、エクサバイトにわたる拡張性

様々な分析および機械学習ツールとの連携

データの移動を伴わずにデータを処理

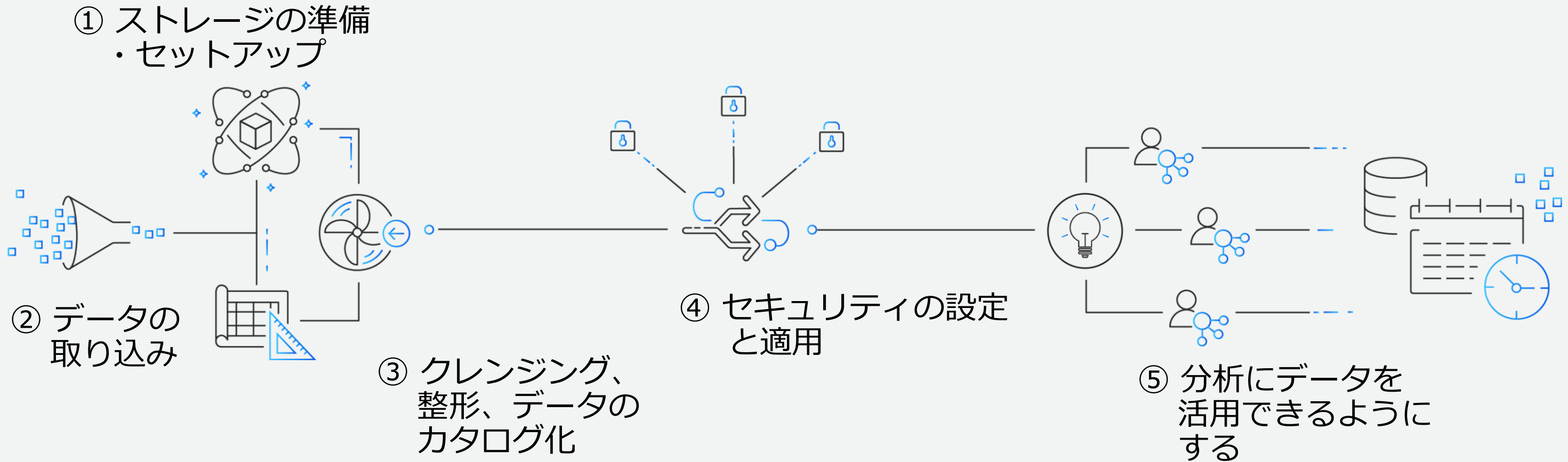
低コストなデータの保存と分析

AWS 上にデータレイク & 分析環境を構築されたお客様



AWS Lake Formation 登場の背景

データレイク構築時の代表的な作業フロー



多くの作業ステップが必要

データレイクの構築には数ヶ月単位の時間が必要

AWS Lake Formation

セキュアなデータレイクを短期間で構築



データの認識、取り込み、
整形、変換

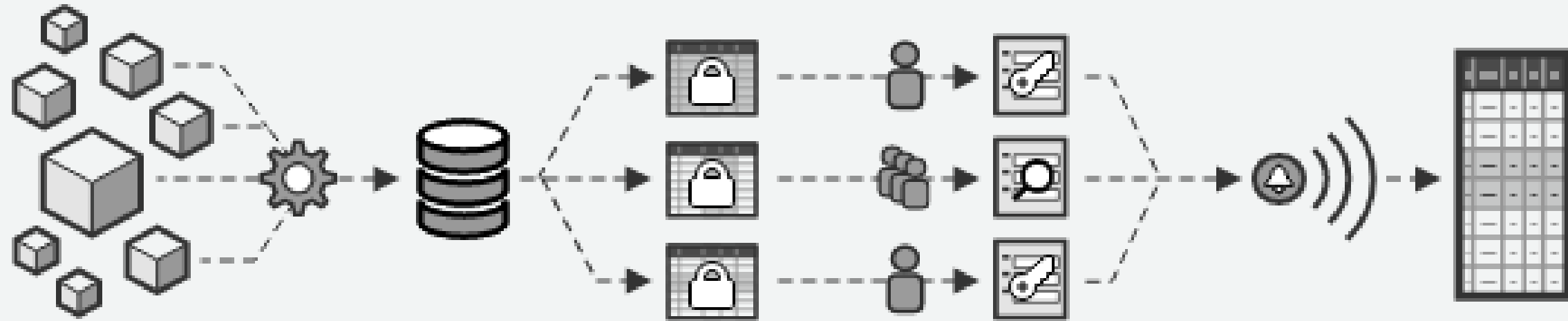


複数サービス間でセキュリティ
ポリシーを適用



新たなインサイトを
提供するカタログ管理

AWS Lake Formation 概要



データ取込みと 構造化

- 自動的にデータ取込み、成形、暗号化して、既存のAmazon S3バケットに登録

セキュリティ & コントロール

- 適切なユーザー、グループに正しいデータへのアクセス制御を定義
- データベース、表、列の単位の粒度で制御可能

協調 & 利用

- メタデータカタログを利用した検索と定義確認
- 全てのアクセスはIAMポリシーによりチェック
- 新しいデータが取込まれたり、ツールが変更されてもポリシーにより保護可能

監視 & 監査

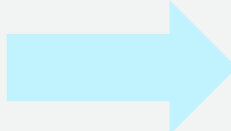
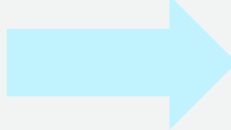
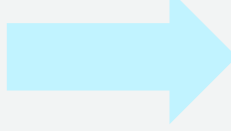
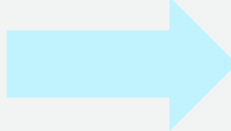
- アクセス要求や発生したポリシー例外を記録
- アクティビティ履歴で詳細に変更ログやデータの入手経路をレビュー

AWS Lake Formation の機能

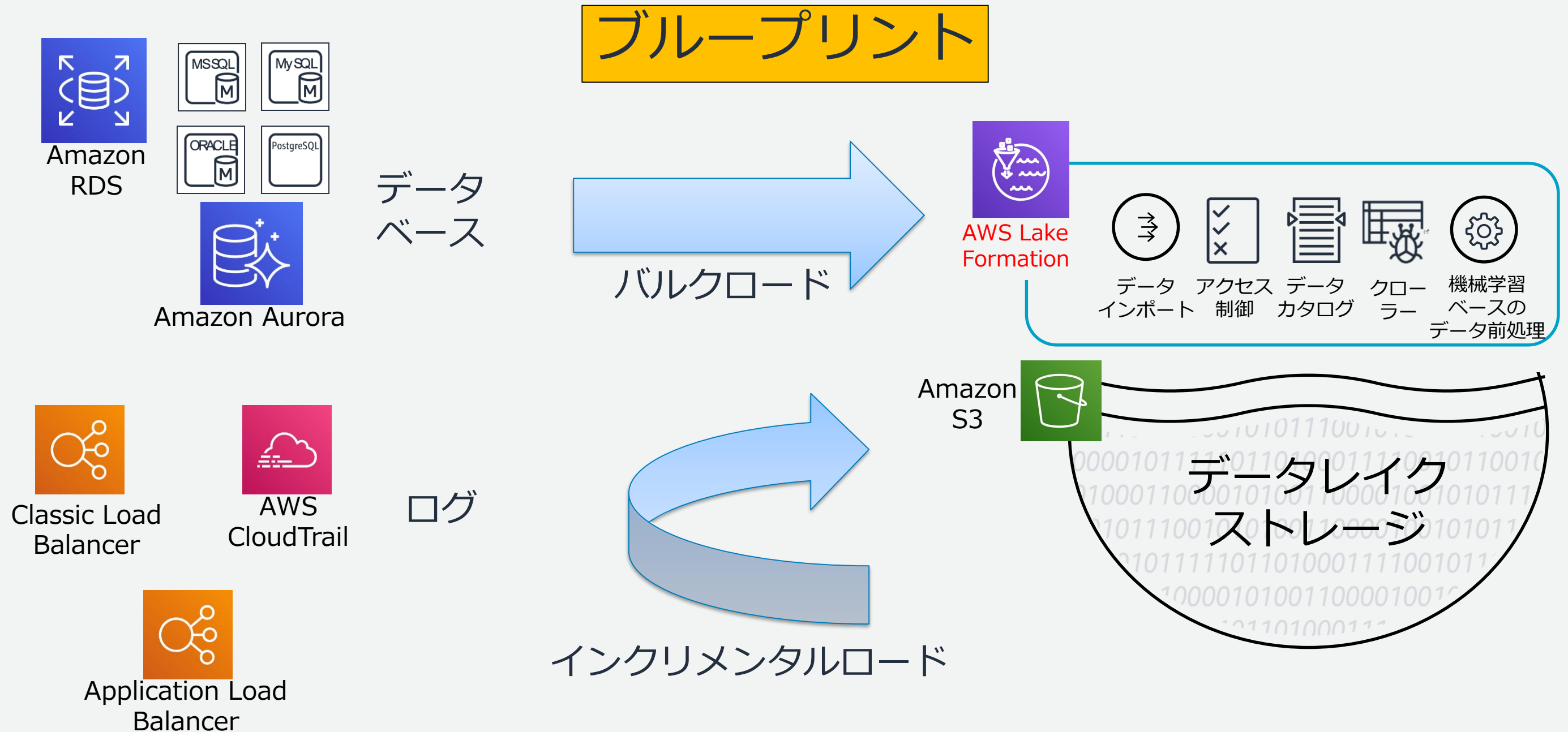
Lake Formation の各コンポーネント



AWS Lake Formation

- データ取込みと構造化  **ブループリント**：汎用的なデータ取込みテンプレートを使い、自動的なデータ取込みと構造化を実現
- セキュリティ & コントロール  **パーミッション**：SQL ライクな Grant/Revoke でシンプルなアクセス制御を実現
- 協調 & 利用  **データカタログ**：スキーマやロケーションなどのデータのメタ情報を管理し、ファセット検索で探したいデータセットを探しやすく
- 監視 & 監査  **ロギング**：コンソールによる直近のアクティビティの詳細を確認

AWS Lake Formation – 主要機能



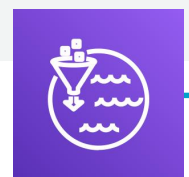
AWS Lake Formation – 主要機能

アクセスパーミッション



管理者

1. Lake Formation 上でユーザーのアクセス制御を設定



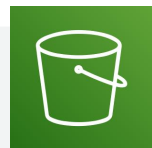
AWS Lake Formation



アクセス制御

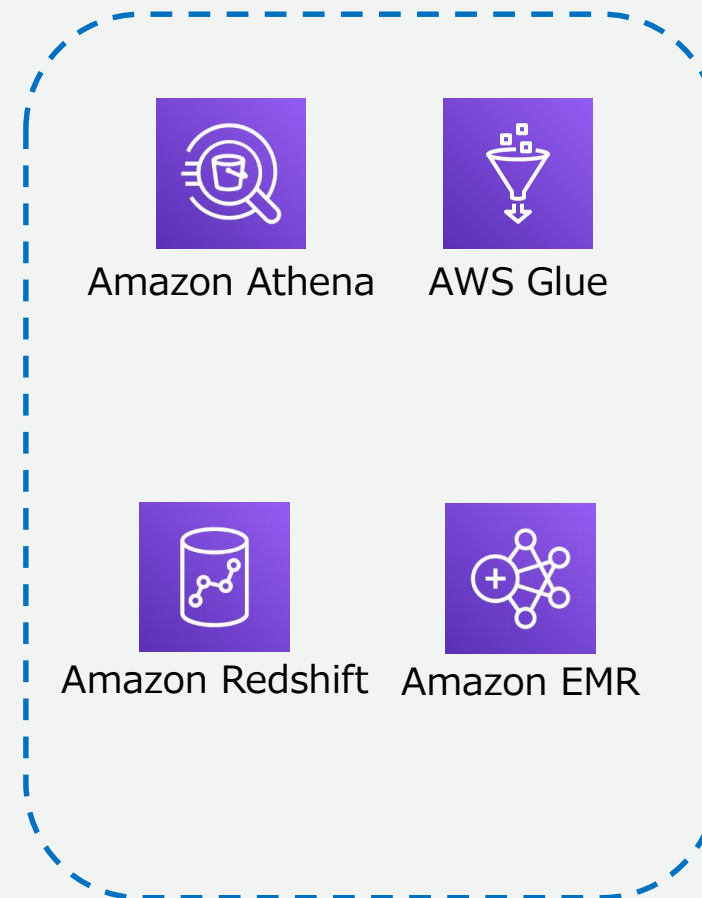


データカタログ



Amazon S3

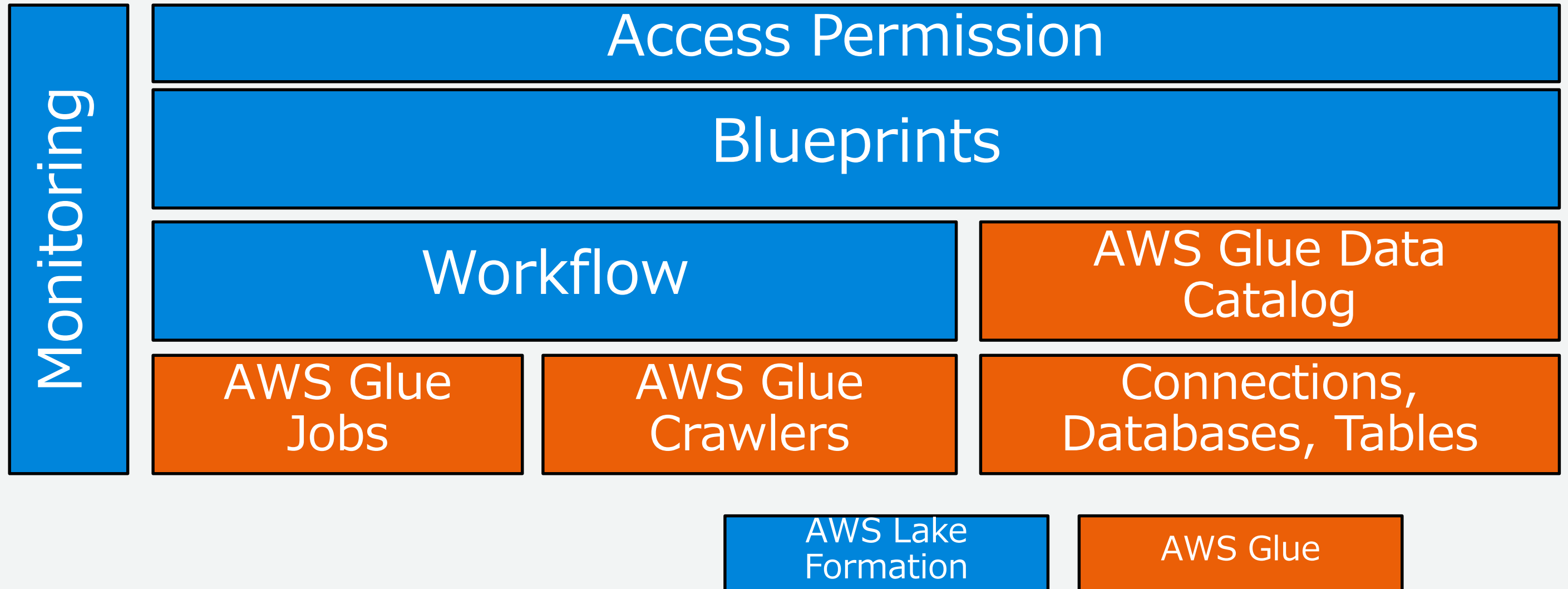
データレイク
ストレージ



2. ユーザーは、任意のサービスからデータへアクセス

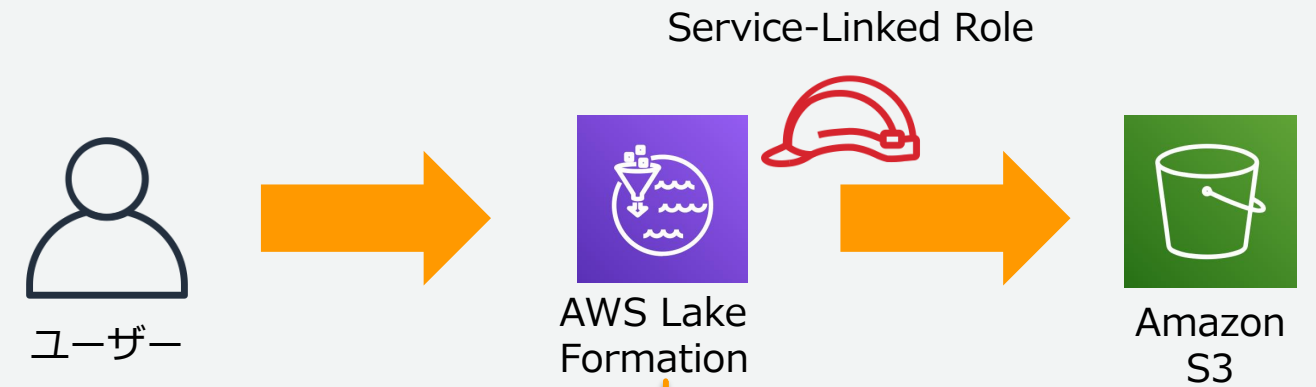
Lake Formation と AWS Glue の関係

Lake Formation は Glue の拡張機能と言え、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などでより便利に Glue の機能を使えるようになっています。



データソースとして S3 を Lake Formation に登録

- S3 は Lake Formation のストレージ層
- データソースとして、既存 S3 パスを Lake Formation に登録
- Lake Formation と信頼関係があり S3 パスへの読み取り/書き込み権限を持つ IAM ロールを選択
- S3 パスに格納されたデータに、Lake Formation からアクセスが可能に



登録することではじめて
Lake Formation での権
限制御の対象となります

データソースとして S3 を Lake Formation に登録

設定は、**Data Lake Location** をクリックし、
[Register location]

S3 パスと Service-Linked Role を指定する

AWS Lake Formation > Data lake locations

Data lake locations (1/7)

Search: tmpuehara

Amazon S3 path	IAM role	Last modified
s3://tmpuehara1	AWSServiceRoleForLakeFormationData...	Mon, Sep 9, 2019, 7:36 AM UTC



Register location

Amazon S3 location
Register an Amazon S3 path as the storage location for your data lake.

Amazon S3 path
Choose an Amazon S3 path for your data lake.

s3://xxx/yyy

Review location permissions - strongly recommended
Registering the selected location may result in your users gaining access to data already at that location. Before registering a location, we recommend that you review existing location permissions on resources in that location.

IAM role
To add or update data, Lake Formation needs read/write access to the chosen Amazon S3 path. Choose a role that you know has permission to do this, or choose the **AWSServiceRoleForLakeFormationDataAccess** service-linked role. When you register the first Amazon S3 path, the service-linked role and a new inline policy are created on your behalf. Lake Formation adds the first path to the inline policy and attaches it to the service-linked role. When you register subsequent paths, Lake Formation adds the path to the existing policy.

AWSServiceRoleForLakeFormationDataAccess

セキュリティ

Lake Formation 用語解説

✓ Data Lake Location

Lake Formation でアクセスパーミッションを行う対象の **S3 パスを登録**します。

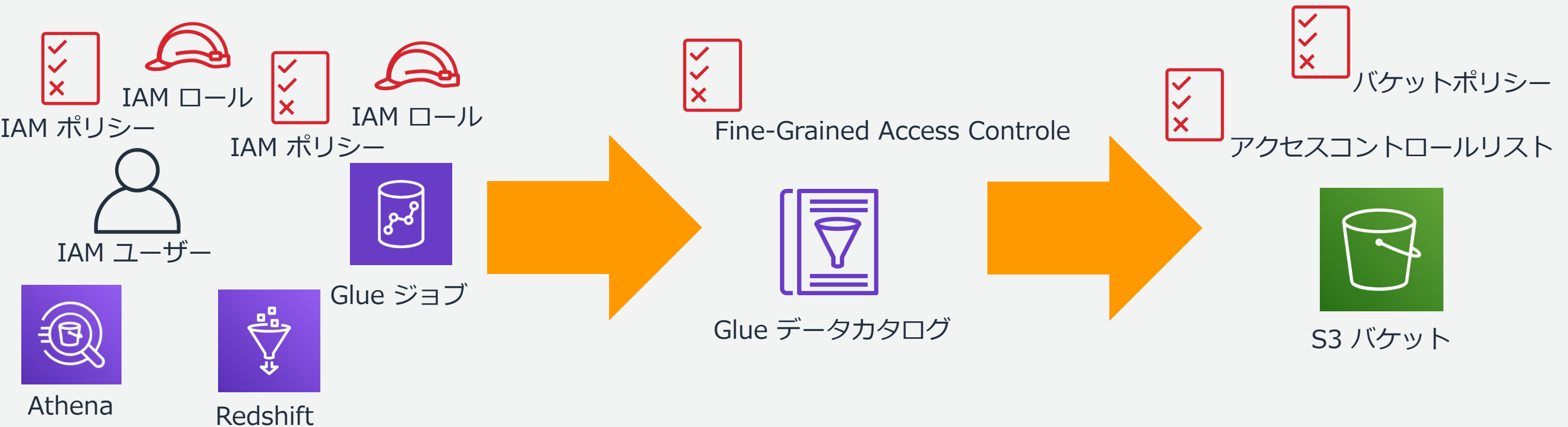
✓ Data Location

プリンシパル※に対して、Data Lake Location で登録した S3 パスにアクセス許可を与えます。これにより**データベース**や**テーブル**の作成ができるようになります。

※注釈：プリンシパルは IAM ユーザーや IAM ロールのこと

Lake Formation 以前のセキュリティ設定

これまで IAM と S3、Glue データカタログを組み合わせて実現していたデータレイクのアクセスパーミッションでは、複数の異なる箇所に別々にポリシーを登録する必要がありました。例えば、IAM ユーザーに特定のバケットへの S3 の権限や特定のデータベースやテーブルへの Glue の権限を与え、S3 バケットポリシーでも特定バケットに対して権限を与えていました。



Lake Formation 以前のセキュリティ

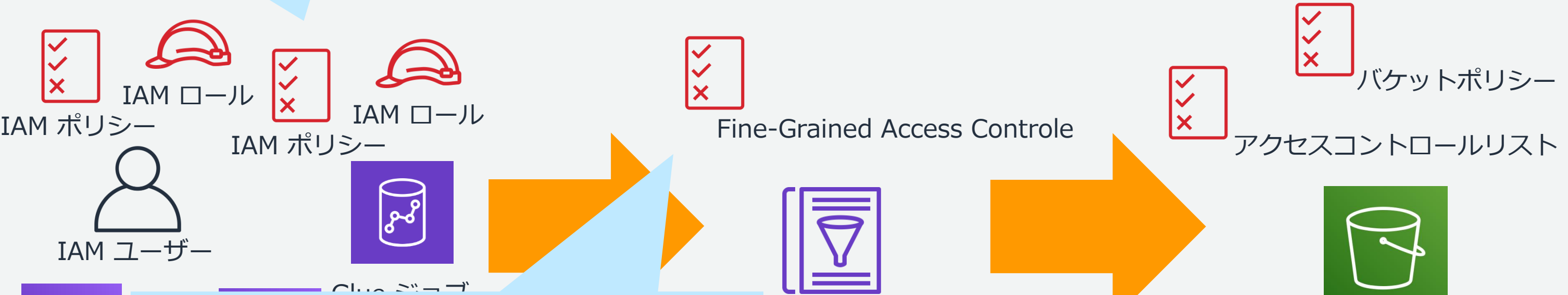
```
{ "Version": "2012-10-17", "Statement": [ { "Sid": "statement1", "Effect": "Allow", "Action": [ "s3:PutAnalyticsConfiguration", "s3:GetObjectVersionTagging", "s3:GetBucketCORS", "s3:GetBucketLocation", "s3:GetObjectVersion", "s3:ReplicateDelete", ... ], "Resource": "arn:aws:s3:::testbucket" }, { "Sid": "statement2", "Effect": "Allow", "Action": [ "s3:GetAccountPublicAccessBlock", "s3:ListAllMyBuckets", "s3:ListJobs", "s3:CreateJob", "s3:HeadBucket", ... ], "Resource": "*" } ] }
```

```
{ "Version": "2012-10-17", "Statement": [ { "Sid": "statement1", "Effect": "Allow", "Principal": { "AWS": "arn:aws:iam::xxxxxxxxxxxx:user/taro" }, "Action": [ "s3:GetBucketLocation", "s3:ListBucket" ], "Resource": [ "arn:aws:s3:::testbucket" ] }, { "Sid": "statement2", "Effect": "Allow", "Principal": { "AWS": "arn:aws:iam::xxxxxxxxxxxx:user/taro" }, "Action": [ "s3:GetObject" ], "Resource": [ "arn:aws:s3:::testbucket/*" ] } ] }
```

ログを結
異なる
定のバ

S3 バケットポリシーでも特... ットに

対して権限... えていました。



Athena

```
{ "Version" : "2012-10-17", "Statement" : [ { "Effect" : "Allow", "Principal" : { "AWS" : "arn:aws:iam::xxxxxxxxxxxx:user/taro" }, "Action" : "glue:*", "Resource" : [ "arn:aws:glue:ap-northeast-1:xxxxxxxxxxxx:catalog", "arn:aws:glue:ap-northeast-1:xxxxxxxxxxxx:database/dba", "arn:aws:glue:ap-northeast-1:xxxxxxxxxxxx:table/dba/tablea" ] } ] }
```

データカタログ



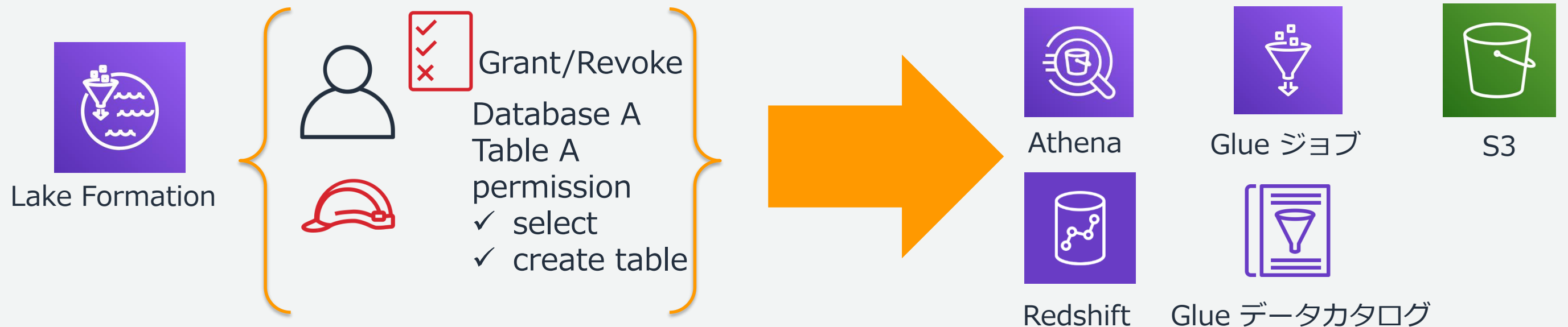
S3 バケット



Lake Formation のセキュリティパーミッション

Lake Formation は、IAM を拡張した独自のアクセス許可モデルを持ち、データレイク内のデータへのアクセスを保護します。

Lake Formation のアクセス許可モデルでは IAM ユーザーや IAM ロールに対して、SQL ライクでシンプルな Grant/Revoke により、データレイクに格納されているデータの一元的なきめ細かいアクセス制御が可能です。



Lake Formation におけるセキュリティパーミッション

Lake Formation では、**データロケーションのアクセス許可**、**データカタログのアクセス許可**、**データアクセス許可**の3種類に明示的に許可を与えます。さらに明示的な許可またはデータベースやテーブル作成の結果として**暗黙的なアクセス許可**が与えられます。

- データロケーションのアクセス許可
- データアクセス許可
- データカタログのアクセス許可
- 暗黙的なアクセス許可

データロケーションのアクセス許可

データロケーションは、データが保存される S3 パスです。登録された S3 のある場所にデータベースまたはテーブルを作成するために、プリンシパルにその場所に対するアクセス許可を与えます。

※ブループリントを使う場合は、S3 パスと、Lake Formation がブループリントにより作成されたワークフローで対象のロケーションに対してテーブルを作成するための IAM ロール を指定します。

Grant 選択

Data locations (7)
Choose a storage location for which to review, grant or revoke user permissions.

	Principal	Principal type	Resource	Grantable
<input type="radio"/>	handson-LakeFormationWorkflowRole	IAM role	s3://handson-uehara-datalake-rdb	-
<input type="radio"/>	handson-LakeFormationWorkflowRole	IAM role	s3://handson-uehara-datalake-trail	-
<input type="radio"/>	uehara	IAM user	s3://uehara-datalake-cloudtrail-v	-
<input type="radio"/>	LakeFormationWorkflowRole	IAM role	s3://uehara-datalake-cloudtrail-vout	-
<input type="radio"/>	LakeFormationWorkflowRole1	IAM role	s3://uehara-datalake-tutorial	-

Grant permissions
Add access permissions for specific storage locations.

IAM users and roles
Add one or more IAM users or roles.

LakeFormationWorkflowRole
Role

Active Directory users and groups (EMR beta only)
Enter one or more Active Directory users or groups.

Storage locations
Choose one or more data lake locations.

s3://uehara-datalake-tutorial

Grantable

Cancel Grant

データカタログのアクセス許可

データカタログには、基となるデータに関するメタデータが格納されます。メタデータはデータベースとテーブルとして編成されています。テーブルは S3 に保存されているデータの場所を指します。データベースはテーブルの集合です。データカタログのアクセス許可は、データベースとテーブルを作成、編集、および削除する権限をシンプルな Grant/Revoke で与えます。(例 CREATE_DATABASE, CREATE_TABLE)

CREATE_TABLEの例

Data permissions (46)
Choose a database or table for which to review, grant or revoke user permissions.

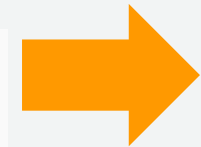
🔍

🔄 Revoke **Grant**

< 1 2 3 ... > ⚙️

	Principal	Principal type	Resource type	Resource	Permissions	Grantable
<input type="radio"/>	LakeFormationWorkflowRole	IAM role	Database	datalake_jdbc	All, Alter, Create table, Drop	-
<input type="radio"/>	uehara	IAM user	Database	datalake_jdbc	Alter, Create table, Drop	Alter, Create, Drop
<input type="radio"/>	handson-LakeFormationWorkflowRole	IAM role	Database	handson-lakeformation_rdb	All, Alter, Create table, Drop	-
<input type="radio"/>	uehara	IAM user	Database	handson-lakeformation_rdb	Alter, Create table, Drop	Alter, Create, Drop
<input type="radio"/>	uehara	IAM user	Database	tmp1	All, Alter, Create table, Drop	All, Alter, Create table, Drop

Grant 選択



Grant permissions ✕

Choose the access permissions to grant. IAM permissions must also allow access.

IAM users and roles
Add one or more IAM users or roles.

Choose IAM principals to add ▼

uehara ✕
User

Active Directory users and groups (EMR beta only)
Enter one or more Active Directory users or groups.

Ex: `arn:aws:iam::<account-id>:saml-provider/ADTestIDP:user/<user>/</user>`

Database
Add one or more databases.

Choose databases ▼

datalake_jdbc ✕

Table - optional
Add one or more tables.

Choose tables ▼

Database permissions
Choose the specific access permissions to grant.

Create table Alter Drop

All

データアクセス許可

テーブルの基になるデータを読み書きするために、プリンシパルにテーブルに対するデータアクセス許可をシンプルな Grant/Revoke で与えます。

(例 SELECT, INSERT)

The screenshot shows the AWS IAM console 'Tables' page. A table named 'datalakejdbc_dblf_person' is selected. The 'Actions' dropdown menu is open, and the 'Permissions' sub-menu is visible. The 'Grant' option is highlighted with a red box. A large orange arrow points from the 'Grant' option to the right, towards the 'Grant permissions' dialog box.

Name	Database	Location	Classification	Last updated
datalakejdbc_dblf_person	datalake_jdbc	s3://uehara-...	...	2019年9月1日(日) 7:37 UTC
_temp_datalakejdbc_dblf_...	datalake_jdbc	s3://uehara-...
test200m_dblf_item	lf01	s3://uehara-...
_temp_test200m_dblf_item	lf01	s3://uehara-...	...	2019年9月1日(日) 1:59 UTC

The screenshot shows the 'Grant permissions' dialog box for the table 'datalakejdbc_dblf_person'. The 'IAM users and roles' section has 'uehara User' selected. The 'Table permissions' section has 'Insert' and 'Drop' checked. The 'Include columns' section is set to 'Include columns'.

Grant permissions datalakejdbc_dblf_person

Grant access permissions to specific users and roles.

IAM users and roles
Add one or more IAM users or roles.
Choose IAM principals to add
uehara User

Active Directory users and groups (EMR beta only)
Enter one or more Active Directory users or groups.
Ex: arn:aws:iam::<account-id>:saml-provider/ADTestIDP:user

Column - optional
Choose filter type
Include columns

Include columns
Grant permissions to access the selected columns.
Choose columns
date string, name string

Table permissions
Choose the specific access permissions to grant.
 Select all Alter Insert Drop
 Delete Select

Grant 選択

データアクセス許可

テーブルおよび列レベルの権限付与

列レベルのアクセス制御の指定

Column - optional

Choose filter type

Include columns ▼

Q |

None

Include columns

Exclude columns

Table permissions

Choose the specific access permissions to grant.

- Select all Alter Insert Drop
- Delete Select



ユーザー 1

一部の列のみ
アクセス可能

Column name	Data type
marketplace	string
customer_id	bigint
review_id	string
product_id	string
product_parent	bigint
product_title	string
star_rating	string
helpful_votes	bigint
total_votes	bigint
vine	string
verified_purchase	string
review_headline	string
review_body	string
review_date	string
product_category	string



ユーザー 2

全ての列に
アクセス可能

暗黙的なアクセス許可

データカタログ権限の明示的な付与は、追加の暗黙的な権限の付与をする場合があります。データベースの作成など特定の Lake Formation タスクを実行すると、暗黙的な権限の付与も行われます。

- データベース作成者
作成するデータベース内のすべてのテーブルに対するすべての権限を持ちます。
- テーブル作成者
作成するテーブルに対するすべての権限を持ちます。作成するテーブルにアクセス許可を付与できます。
- データレイクユーザー
権限を持つデータベースまたはテーブルを表示および一覧表示できます。

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

データレイク管理者の暗黙的なアクセス許可

Lake Formation では、**データレイク管理者**※に以下の**暗黙的なアクセス許可**が与えられます

- データカタログ内のすべてのオブジェクトへの完全なメタデータアクセスがあります。このアクセスは管理者から取り消すことはできません。
- データレイク内のすべての Data Location Permission があります。
- データカタログ内のオブジェクトへのアクセスをプリンシパルに許可できます。このアクセスは管理者から取り消すことはできません。
- データカタログにデータベースを作成できます。
- 別のユーザーにデータベースを作成する権限を付与できます。
- データロケーションに S3 パスを登録できます。

データレイク管理者には、データベースを削除したり、テーブルを変更または削除するための暗黙的な権限はありません。ただしそのための権限を自分で付与できます。

※データレイク管理者は IAM の Administrator 権限を持つユーザーではなく、Lake Formation で定義される管理者です。

アクセス許可の一覧

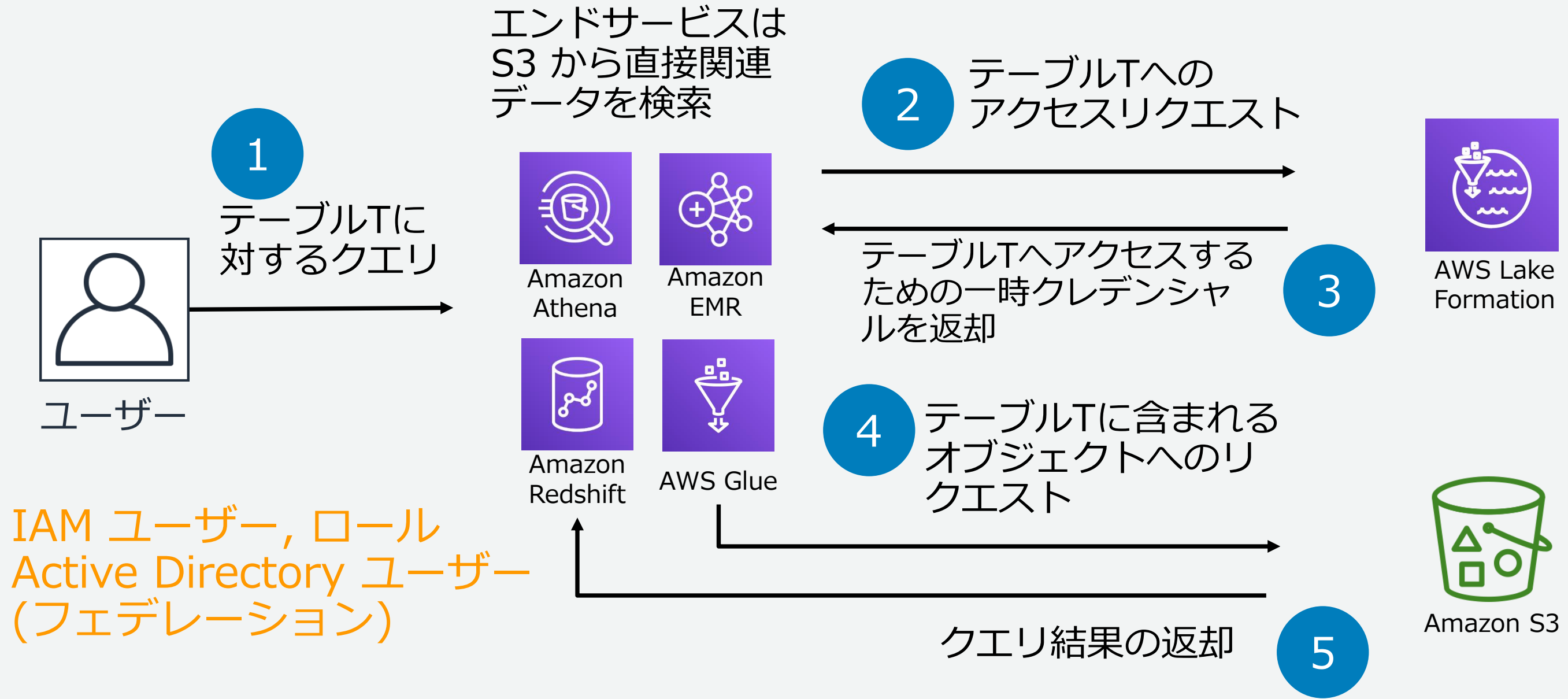
許可の種類	CREATE DATABASE, CREATE TABLE (登録ロ ケーションを指定) ※1	CREATE DATABASE, CREATE TABLE ※1	SELECT, INSERT	主に作成したリソース に対する権限
データロケーション のアクセス許可	●			
データカタログの アクセス許可	●	●		▲ ※2
データアクセス許 可			●	▲ ※2
暗黙のアクセス許 可				●

※1 CREATE_DATABASE や CREATE_TABLE は留意すべきケースがいくつかあるので詳細は以下のリンクを確認ください。

※2 特定のデータベースや特定のテーブルへのアクセス権限

https://docs.aws.amazon.com/ja_jp/lake-formation/latest/dg/data-catalog-permissions.html

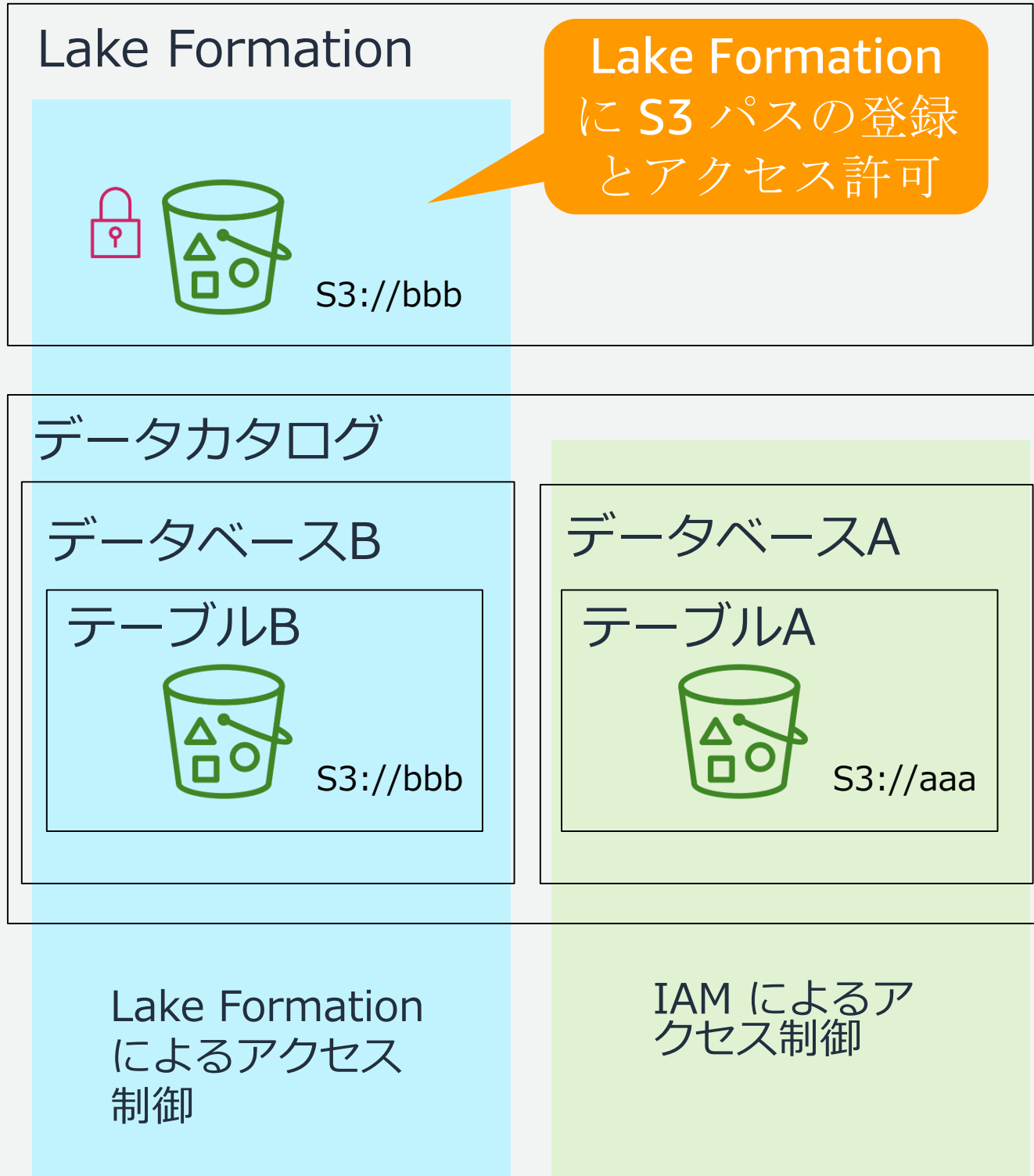
アクセス時の連携イメージ



従来の Glue によるアクセス制御との違い

データベースAは Lake Formation に登録されていない S3 パスのテーブルがあるのでデータカタログのアクセス許可と S3 アクセス許可などの IAM ポリシーにより制御されます。これは従来どおりです。

データベースBは Lake Formation に登録とアクセス許可されている S3 パスのテーブルがあるので Lake Formation の権限により制御されます。



Lake Formation のペルソナ[※]と権限のリファレンス

ペルソナ	説明
IAM 管理者(AWS管理者)	少なくとも IAM と S3 へのフル権限を持つ。または AdministratorAccess のIAM ポリシーを持つ。
データレイク管理者	データカタログへアクセスし、データベースを作成し、他のユーザーに Lake Formation による権限付与が行える。IAM 管理者より持っている IAM アクセス許可は少ないが、データレイクを管理するのに十分な権限を持つ。
データエンジニア (オプション)	所有するデータベースへの権限を持ち、ブループリントからワークフローを作成し、ワークフローを実行するのに十分な権限を持つ。
データ分析者 (オプション)	Athena などを使用して、データレイクに対してクエリを実行できるユーザー。クエリを実行するための権限のみを持つ。
ワークフローロール	ユーザーに代わってワークフローを実行するロール。ブループリントからワークフローを作成するときに、このロールを指定します。

※ペルソナ：データレイクを使うユーザータイプ

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

Lake Formation のペルソナ※と権限のリファレンス

ペルソナ	IAM, S3 全権限	Lake Formation の全権限	Lake Formation の一部の権限	Pass Role 権限	Athena 等の分析サービスの権限
IAM 管理者(AWS 管理者)	●				
データレイク管理者		●		●	
データエンジニア (オプション)			●	●	
データ分析者 (オプション)			●		●
ワークフローロール			●	●	

※ペルソナ：データレイクを使うユーザータイプ

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

データレイク管理者の権限

ポリシータイプ	ポリシー
AWSマネージドルール	AWSGlueConsoleFullAccess
インラインポリシー (basic)	<pre>{ "Version": "2012-10-17", "Statement": [{ "Effect": "Allow", "Action": ["lakeformation:*", "cloudtrail:DescribeTrails", "cloudtrail:LookupEvents", "iam:PutRolePolicy", "iam:CreateServiceLinkedRole"], "Resource": "*" }, { "Effect": "Deny", "Action": ["lakeformation:PutDataLakeSettings"], "Resource": "*" }] }</pre>
インラインポリシー (pass role)	<pre>{ "Version": "2012-10-17", "Statement": [{ "Sid": "PassRolePermissions", "Effect": "Allow", "Action": ["iam:PassRole"], "Resource": ["arn:aws:iam::<i>account-id</i>:role/<i>workflow_role</i>"] }] }</pre>

データエンジニアの権限

ポリシータイプ	ポリシー
AWSマネージドルール	AWSGlueServiceRole
インラインポリシー (data access)	<pre>{ "Version": "2012-10-17", "Statement": [{ "Sid": "Lakeformation", "Effect": "Allow", "Action": ["lakeformation:GetDataAccess", "lakeformation:GrantPermissions"], "Resource": "*" }] }</pre>
インラインポリシー (pass role)	<pre>{ "Version": "2012-10-17", "Statement": [{ "Sid": "PassRolePermissions", "Effect": "Allow", "Action": ["iam:PassRole"], "Resource": ["arn:aws:iam::<i>account-id</i>:role/<i>workflow_role</i>"] }] }</pre>
インラインポリシー (for ingesting data outside the data lake, for example Cloudtrail logs)	<pre>{ "Version": "2012-10-17", "Statement": [{ "Effect": "Allow", "Action": ["s3:GetObject", "s3:ListBucket"], "Resource": ["arn:aws:s3::<i>your-s3-cloudtrail-bucket</i>/*"] }] }</pre> <p>https://docs.aws.amazon.com/ja_jp/lake-formation/latest/dg/permissions-reference.html</p>

データ分析者の権限

ポリシータイプ	ポリシー
AWSマネージドルール	AmazonAthenaFullAccess
インラインポリシー (basic)	<pre>{ "Version": "2012-10-17", "Statement": [{ "Effect": "Allow", "Action": ["lakeformation:GetDataAccess", "glue:GetTable", "glue:GetTables", "glue:SearchTables", "glue:GetDatabase", "glue:GetDatabases", "glue:GetPartitions"], "Resource": "*" }] }</pre>

ワークフローロールの権限

ポリシータイプ	ポリシー
AWSマネージドルール	AWSGlueServiceRole
インラインポリシー (data access)	<pre>{ "Version": "2012-10-17", "Statement": [{ "Sid": "Lakeformation", "Effect": "Allow", "Action": ["lakeformation:GetDataAccess", "lakeformation:GrantPermissions"], "Resource": "*" }] }</pre>
インラインポリシー (pass role)	<pre>{ "Version": "2012-10-17", "Statement": [{ "Sid": "PassRolePermissions", "Effect": "Allow", "Action": ["iam:PassRole"], "Resource": ["arn:aws:iam::<i>account-id</i>:role/<i>workflow_role</i>"] }] }</pre>
インラインポリシー (for ingesting the data lake, for example CloudTrail logs)	<pre>{ "Version": "2012-10-17", "Statement": [{ "Effect": "Allow", "Action": ["s3:GetObject", "s3:ListBucket"], "Resource": ["arn:aws:s3::<i>your-s3-cloudtrail-bucket</i>/*"] }] }</pre> <p>https://docs.aws.amazon.com/ja_jp/lake-formation/latest/dg/permissions-reference.html</p>

許可を与えるシナリオの例

シャーリーはデータレイク管理者です。彼女は、会社である AnyCompany のデータレイクをセットアップしたいと考えています。現在、すべてのデータは S3 に保存されています。ジョンはマーケティングマネージャーで、顧客の購入情報(s3://customerPurchases)への書き込みアクセスが必要です。この夏マーケティングアナリストのディエゴが加わります。ジョンには、シャーリーが関与することなくデータに対してクエリを実行するためのディエゴへのアクセスを許可することが必要です

ペルソナ	担当者
データレイク管理者	シャーリー
データエンジニア	ジョン
データ分析者	ディエゴ

https://docs.aws.amazon.com/ja_jp/lake-formation/latest/dg/security-permissions-example-scenario.html

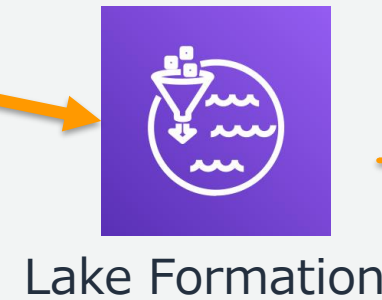
許可を与えるシナリオの例

データレイク管理者



①

① 顧客の購入情報を含む S3 パスを Lake Formation に登録します。



登録

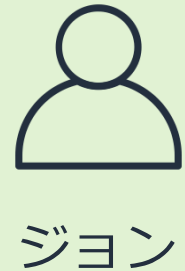
S3://xxx/yyy/

許可

```
GrantPermissions(John,  
S3Location("s3://xxx/yyy/"),  
[DATA_LOCATION_ACCESS])
```

```
GrantPermissions(John, catalog,  
[CREATE_DATABASE])
```

データエンジニア



② ③



② 顧客の購入情報を含む S3 パスへのアクセスをジョンに許可します。



③ データベースを作成する許可をジョンに与えます。

データ分析者

許可を与えるシナリオの例

① データベース John_DB を作成します。ジョンはデータベースを作成したため、自動的にそのデータベースに対する CREATE_TABLE 権限を持ちます(暗黙的な許可)。

データレイク管理者

データエンジニア

データ分析者



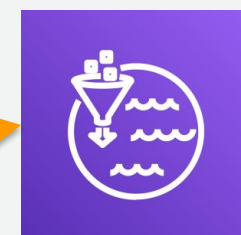
ジョン



ディエゴ

① ②

③



Lake Formation

データベース
John_DB

テーブル
John_Table

許可

GrantPermissions(Diego, John_Table,
[SELECT])

② s3://xxx/yyy/ を指すテーブル John_Table を作成します。ジョンはテーブルを作成したため、このテーブルに対する全ての権限を持ちます(暗黙的な許可)。

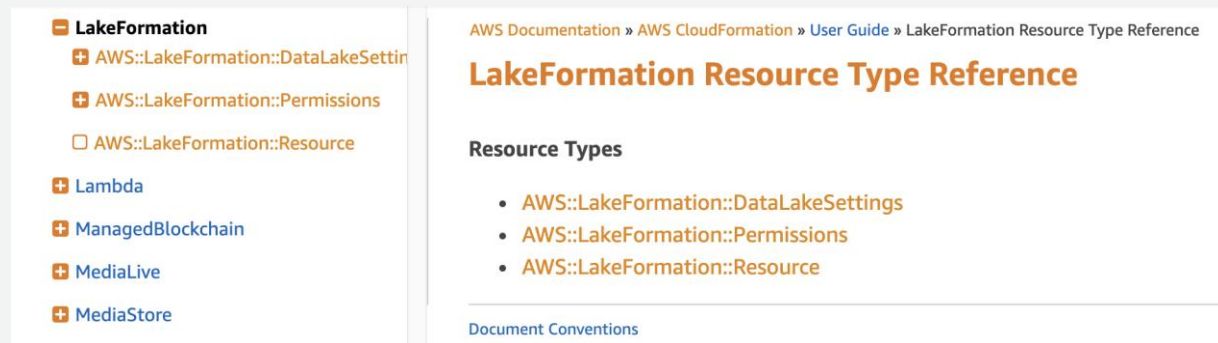


③ アナリストのディエゴに、テーブル John_Table にアクセスする許可を与えます。

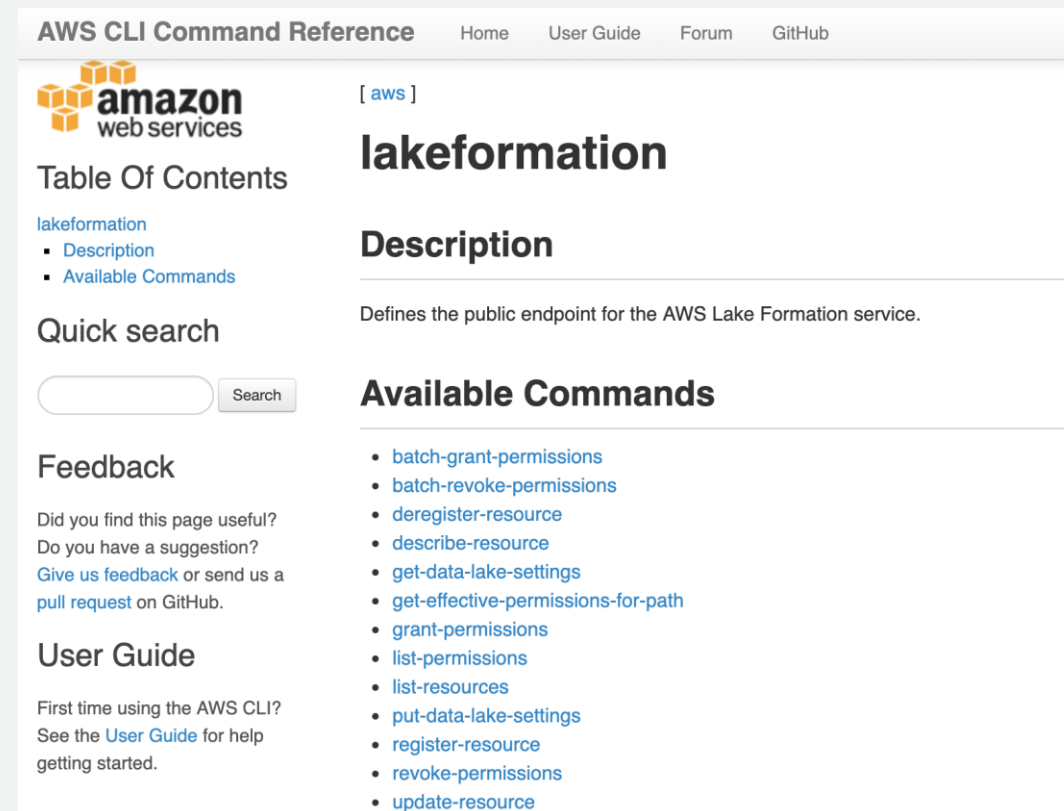
自動化

CloudFormation や AWS CLI で Lake Formation のアクセス許可操作を自動化できます。

- CloudFormation



- AWS CLI



CloudFormation では以下のようなテンプレートでスタックを作成することで、Lake Formation の管理者設定が可能です。

```
{
  "AWSTemplateFormatVersion": "2010-09-09",
  "Resources": {
    "LFDLS4VYBJ": {
      "Type": "AWS::LakeFormation::DataLakeSettings",
      "Properties": {
        "Admins": [{
          "DataLakePrincipalIdentifier": "arn:aws:iam::xxxxxxxxxxxx:user/uehara"
        }]
      }
    }
  }
}
```



自動化

- AWS CLI では以下のように管理者設定の反映が可能です。

```
$ aws lakeformation put-data-lake-settings --cli-input-json file://ladmin.json
```

- AWS CLI では以下のように管理者設定の内容をgetすることが可能です。

```
$ aws lakeformation get-data-lake-settings
{
  "DataLakeSettings": {
    "DataLakeAdmins": [
      {
        "DataLakePrincipalIdentifier": "arn:aws:iam::xxxxxxxxxxxx:user/uehara"
      }
    ],
    "CreateDatabaseDefaultPermissions": [
      {
        "Principal": {
          "DataLakePrincipalIdentifier": "EVERYONE"
        },
        "Permissions": [
          "ALL"
        ]
      }
    ],
    "CreateTableDefaultPermissions": [
      {
        "Principal": {
          "DataLakePrincipalIdentifier": "EVERYONE"
        },
        "Permissions": [
          "ALL"
        ]
      }
    ]
  }
}
```

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

ladmin.json

```
{
  "DataLakeSettings": {
    "DataLakeAdmins": [
      {
        "DataLakePrincipalIdentifier":
"arn:aws:iam::xxxxxxxxxxxx:user/uehara"
      }
    ],
    "CreateDatabaseDefaultPermissions": [
      {
        "Principal": {
          "DataLakePrincipalIdentifier": "EVERYONE"
        },
        "Permissions": [
          "ALL"
        ]
      }
    ],
    "CreateTableDefaultPermissions": [
      {
        "Principal": {
          "DataLakePrincipalIdentifier": "EVERYONE"
        },
        "Permissions": [
          "ALL"
        ]
      }
    ]
  }
}
```



ニアリアルタイム監査とモニタリング

CloudTrail を有効化しておくことで、Lake Formation のコンソールでデータアクセスを監査できます。[View Event] でどのユーザーまたはロールがどのデータにどのサービスでアクセスしたかなどの詳細を確認できます。

The screenshot shows the AWS Lake Formation console interface. On the left is a navigation sidebar with categories like 'Data catalog', 'Register and ingest', and 'Permissions'. The main content area is titled 'Recent access activity (0/50)' and includes a 'View event' button highlighted with a red box. Below this is a table of access events, also highlighted with a red box, showing details such as event name, principal, and alert time.

	Event name	Principal	Alert time
<input checked="" type="radio"/>	ListPermissions	uehara	2019年8月27日(火) 13:15 UTC
<input type="radio"/>	ListPermissions	uehara	2019年8月27日(火) 13:15 UTC
<input type="radio"/>	ListPermissions	uehara	2019年8月27日(火) 13:15 UTC
<input type="radio"/>	ListPermissions	uehara	2019年8月27日(火) 13:15 UTC
<input type="radio"/>	GetDataLakeSettings	uehara	2019年8月27日(火) 13:15 UTC
<input type="radio"/>	GetDataLakeSettings	uehara	2019年8月27日(火) 13:15 UTC

データカタログ機能

Lake Formation データカタログ

Apache Hive メタストアで行うのと同じ方法で、AWS でメタデータを保存、注釈付け、共有できるマネージドサービスです。

1. データカタログ

- Glue データカタログと統合
- Apache Hive メタストア互換

2. 同一 AWS アカウント内で共有可能な統合リポジトリ

3. データベースはテーブルの集合体

4. テーブルはデータのスキーマ情報、ロケーションなどのメタデータを保管

データカタログ

データベースA

テーブル1

- テーブル情報
 □ロケーション : s3://xxx/yyy
- テーブルプロパティ
 env:research
- テーブルスキーマ
- パーティション

テーブル2

...

データベースB

テーブル3

...

メタデータの構成例

テーブル情報

AWS Lake Formation > Tables > se2_out1

se2_out1 Version 0 (Current version) ▼

Table details

Table name
se2_out1

Description
-

Database
[se2](#)

Classification
parquet

Location
<s3://test-glue00/se2/out1/>

Connection
-

Last updated
2019年9月6日(金) 9:17 UTC

テーブルプロパティ

Serde parameters
serialization.format 1

Table properties

CrawlerSchemaDeserializerVersion 1.0 CrawlerSchemaSerializerVersion 1.0 UPDATED_BY_CRAWLER se2_out1 averageRecordSize 194

classification parquet compressionType none

exclusions ["s3://test-glue00/se2/out1/_common_metadata","s3://test-glue00/se2/out1/_metadata"] objectCount 71 recordCount 38

sizeKey 34086 typeOfData file

テーブルスキーマ

Schema Edit schema

Filter Columns

Column #	Name	Data type	Partition key	Comment
1	deviceid	string	-	-
2	uuid	bigint	-	-
3	appid	bigint	-	-
4	country	string	-	-
4	year	string	1	-
5	month	string	2	-
6	day	string	3	-
7	hour	string	4	-

テーブルパーティション

Partitions Filter Partitions

year	month	day	hour	Files	Properties
2017	11	29	15	View files	View properties
2017	12	17	18	View files	View properties
2017	12	14	14	View files	View properties
2017	12	19	14	View files	View properties

データカタログの検索

Lake Formation では、テキストベースのファセット検索を全メタデータに対して行えるため、分析に利用可能なデータセットのカタログにセルフサービスでアクセスできます。

- Resource Attributes 検索
Classification(例えばPARQUET),
Database(例えばDatabaseA), Location(例
例えばs3),Name(テーブル名で例えばTableA)
- キーワード検索
データベース名、テーブル名、列名、
Description などのメタデータをキーワード
検索
- 複数のフィルタ検索
上記のいくつかを組み合わせると検索

※複数のキーワードは組み合わせられない

✓ Classification を PARQUETで検索

The screenshot shows the Lake Formation console interface. At the top, there's a search bar with the filter "Classification : PARQUET" applied. Below the search bar, a table lists several tables. The first three rows are highlighted, showing their names, databases, locations, classifications, and last updated dates.

Name	Database	Location	Classification	Last updated
datalakejdbc_dblf_person	datalake_jdbc	s3://uehara-datalake-tutorial/...	PARQUET	2019年9月1日(日) 7:37 UTC
_temp_datalakejdbc_dblf_...	datalake_jdbc	s3://uehara-datalake-tutorial/...	PARQUET	2019年9月1日(日) 7:37 UTC
test200m_dblf_item	lf01	s3://uehara-datalake-tutorial-...	PARQUET	2019年9月1日(日) 7:37 UTC

✓ Classification を PARQUET、Location を s3、Keyword を
accountnumber(列名) で複数フィルタ検索

The screenshot shows the Lake Formation console interface with multiple filters applied: "Classification : PARQUET", "Location : s3", and "keyword : accountnumber". Below the filters, a table lists the resulting tables.

Name	Database	Location	Classific...	Last updated
datalakejdbc_dblf_person	datalake_jdbc	s3://uehara-datalake-tutorial/...	PARQUET	2019年9月1日(日) 7:37 UTC

データカタログの検索

ビジネス固有の属性をテーブルプロパティやカラムプロパティとして追加できます。

タグを設定

- テーブル `datalakejdbc_dblf_person` のテーブルプロパティに `env:research` を追加
- テーブル `datalakejdbc_dblf_person` のカラム `storeid` とテーブル `test_person` のカラム `storecode` に、カラムプロパティ `storeidenv:dev` を追加

- ✓ テーブルプロパティ (テーブル `datalakejdbc_dblf_person`)

Table properties

CreatedByJob	datalakejdbc_etl_3_175b1974	CreatedByJobRun	jr_7b76a21487f639f1262014a989b4e6bca1bdbc8a787aea6988849e127dc51b90						
LastTransformCompletedOn	2019-09-01 04:08:07.075201	LastUpdatedByJob	datalakejdbc_etl_3_175b1974						
LastUpdatedByJobRun	jr_2f2eb28235fe510980dacf37614941b7981951e00cc2f1519068b7342dd2c87d	SourceConnection	lf-3-connect						
SourceTableName	dblf_person	SourceType	JDBC	TableVersion	2	TransformTime	0:01:17.846746	classification	PARQUET
env	research								

- ✓ カラムプロパティ (テーブル `datalakejdbc_dblf_person` の `storeid`、テーブル `test_person` の `storecode`)

Column name: storeid, Data type: string, Column #: 4, Partition Key:

Column name: storecode, Data type: string, Column #: 6, Partition Key:

Key	Value
storeidenv	dev

データカタログの検索

追加したビジネス固有の属性(テーブルプロパティやカラムプロパティ)で検索できます。マーケティングやリサーチャーなどデータを扱いたい人が欲しいデータを探しやすくなります。

タグによる検索

- テーブルプロパティで“env:research”で検索。
datalakejdbc_dblf_person がヒット



- カラムプロパティで“storeidenv:dev”による検索。
test_person, datalakejdbc_dblf_person の2つのテーブルがヒット



- ✓ テーブル datalakejdbc_dblf_person (テーブルプロパティ env:research)

Tables (1)

Filter by tags and attributes or search by keyword

keyword : env:research X

Name	Database	Location	Classifica...	Last updated
datalakejdbc_dblf_person	datalake_jdbc	s3://uehara-datalake-tut...	PARQUET	2019年9月8日(日) 12:44 UTC

- ✓ テーブル datalakejdbc_dblf_person (storeidカラムプロパティ storeidenv:dev)
- ✓ テーブル test_person (storecodeカラムプロパティ storeidenv:dev)

Tables (2)

Filter by tags and attributes or search by keyword

keyword : storeidenv:dev X

Name	Database	Location	Classifica...	Last updated
datalakejdbc_dblf_person	datalake_jdbc	s3://uehara-datalake-tut...	PARQUET	2019年9月8日(日) 12:44 UTC
test_person	lf01	s3://uehara-datalake-tut...	parquet	2019年9月8日(日) 12:43 UTC

データカタログのアクセス許可モデルの移行

Lake Formation により広範な IAM を使ったアクセス制御を簡単に設定することが出来ます。これに対して Glue のアクセス許可モデルは IAM を介してアクセスを許可します。

従来から Glue をご利用頂いているお客様は Lake Formation のアクセス許可モデルへの移行が必要になります。

IAMAllowedPrincipals

Lake Formation は、既存の Glue データカタログとの互換性のために有効化された“grant All to IAMAllowedPrincipals”で利用を開始します。これには、既存の Glue アクセス許可を引き続き有効にするデフォルトのアクセス許可が含まれています。これは「All to IAMAllowedPrincipals」と呼ばれる特別な許可を介して行われます。

Lake Formation 権限の使用に移行した後、この設定を無効にすることをお勧めします。

Grant permissions

Choose the access permissions to grant. IAM permissions must also allow access.

IAM users and roles

Add one or more IAM users or roles.

Choose IAM principals to add

iam

Group

IAMAllowedPrincipals

Group

Database

Add one or more databases.

Choose databases

Database permissions

Choose the specific access permissions to grant.

Create table Alter Drop

Data permissions (1)

Choose a database or table for which to review, grant or revoke user permissions.

Q

Database: default

Clear filter

Principal

Principal type

Resource type

Resource

Permissions

Grantable

IAMAllowedPrincipals

Group

Database

default

All

-

データカタログのアクセス許可モデルの移行

公式ドキュメントによる手順

https://docs.aws.amazon.com/ja_jp/lake-formation/latest/dg/upgrade-glue-lake-formation.html

- Step1 : 既存のアクセス許可を確認
- Step2 : Lake Formation でこれらの権限を作成
- Step3 : IAM ユーザーに Lake Formation のアクセス許可を付与
- Step4 : データストアを Lake Formation アクセス許可モデルに切り替える
- Step5 : IAM ユーザーに新しい追加の IAM ポリシーを付与
- Step6 : IAM ポリシーをクリーンアップ

ブループリント機能

ブループリントとは

ブループリントは、データレイクにデータを簡単に取込むことを実現する「データ管理のテンプレート」です。汎用的なユースケースのブループリントを複数用意しています。

2種類のブループリント (2019/10/01時点) ※ブループリントは随時追加予定

- データベース用ブループリント
 - MySQL、PostgreSQL、Oracle、SQL Serverからのデータの取込み
- ログファイル用ブループリント
 - 以下のAWS主要ログの取込み
 - AWS CloudTrail
 - Application Load Balancer
 - Classic Load Balancer

Blueprint type
Configure a blueprint to create a workflow.

<input checked="" type="radio"/> Database snapshot Bulk load data to your data lake from MySQL, PostgreSQL, Oracle, and Microsoft SQL Server databases.	<input type="radio"/> Incremental database Load new data to your data lake from MySQL, PostgreSQL, Oracle, and SQL Server databases.	<input type="radio"/> AWS CloudTrail Bulk load data from AWS CloudTrail sources.
<input type="radio"/> Classic Load Balancer logs Load data from Classic Load Balancer logs.	<input type="radio"/> Application Load Balancer logs Load data from Application Load Balancer logs.	

バルクロード or 増分ロード

ブループリントとは

ブループリント上で必要なパラメータを入力し Lake Formation のワークフローを作成することで、Glue のトリガー、ワークフロー、クローラー、ジョブを自動で生成します。

“Database snapshot” の例

Import source
Configure the workflow source.

Database connection
Choose the connection to the data source. [Create a connection in AWS Glue](#)

datalake-tutorial

Source data path
Enter the path from which to ingest data. For JDBC databases with schema support, enter database/schema/table. Substitute the percent (%) wildcard for schema or table.

dblf/person

Exclude patterns - optional
Specify a glob pattern to limit the workflow. Tables that match this pattern will be excluded. [Learn more](#)

Exclude pattern

Enter an exclude pattern

Remove exclusion

Add exclusion

© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

“AWS CloudTrail” の例

Import source
Configure the workflow source.

CloudTrail name
Choose a CloudTrail source.

handson-uehara-trail

Start date
Choose a CloudTrail source start date.

Choose a start date

2019年9月						
日	月	火	水	木	金	土
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	1	2	3	4	5

Target storage location
Choose a data lake location or other Amazon S3 path.

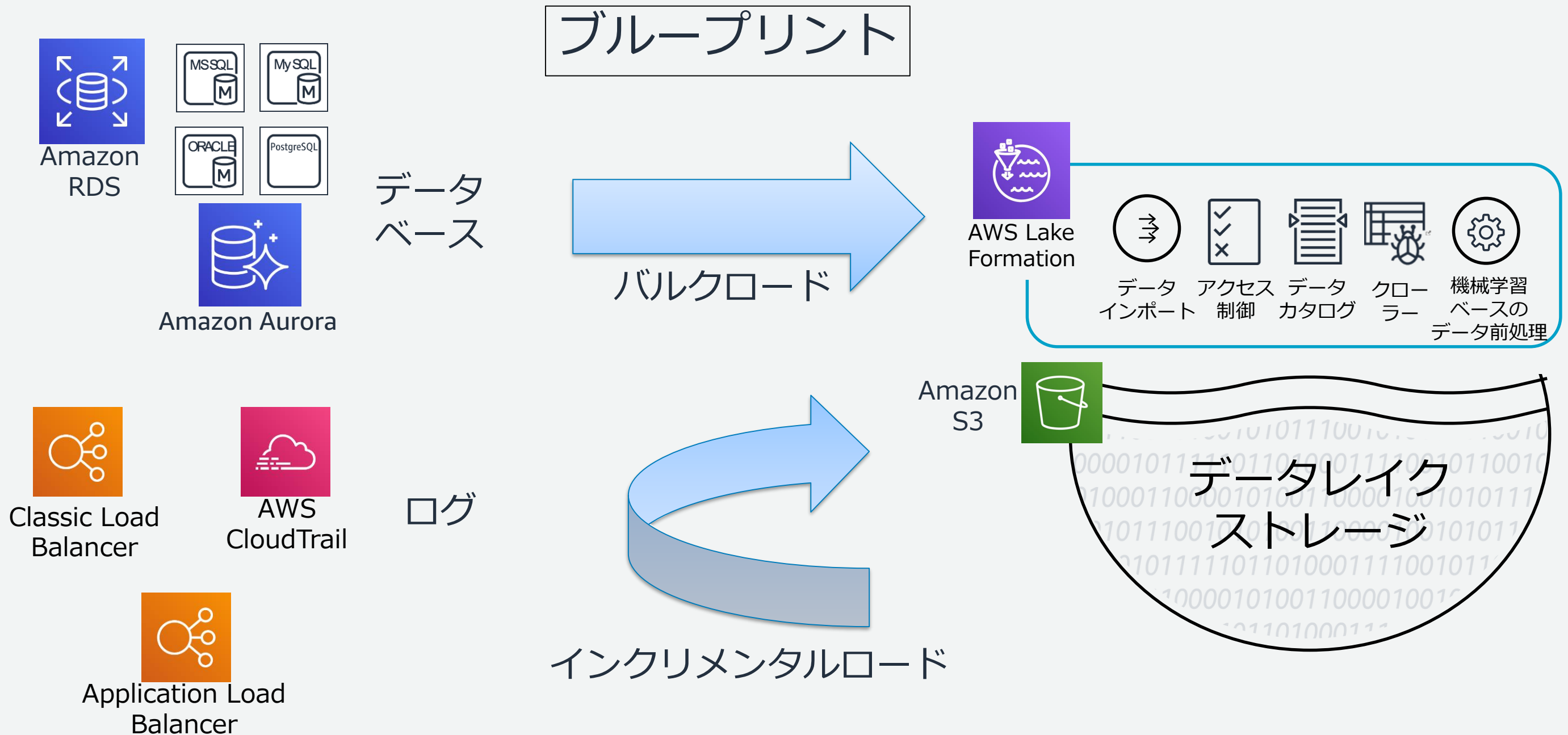
s3://uehara-datalake-tutorial

Browse

Data format
Choose the output data format.

Parquet

ブループリントによるデータレイクへの容易なデータロード



ワークフロー：ブループリントから作成される処理フロー

ブループリントを使い作成された Lake Formation のワークフローは、[Start] で実行することができます。ワークフローの画面では実行履歴の確認や [View graph] でワークフローの詳細を確認することができます。

AWS Lake Formation > Blueprints > datalakejdbc

datalakejdbc

Start

Delete

View graph

Workflow details

Name

datalakejdbc

Last run status

✔ COMPLETED

Last updated

2019年8月24日(土) 8:07 UTC

Created on

2019年8月24日(土) 8:07 UTC

Workflow runs (3)

🔍 Filter Workflow runs

< 1 >



Name

Started on

Run ID

datalakejdbc

2019年9月1日(日) 3:24 UTC

wr_4733c2eb4c172f604efff501344cb0aa95d9786ce3c9aaa9d6bf86b361fe243f [🔗](#)

datalakejdbc

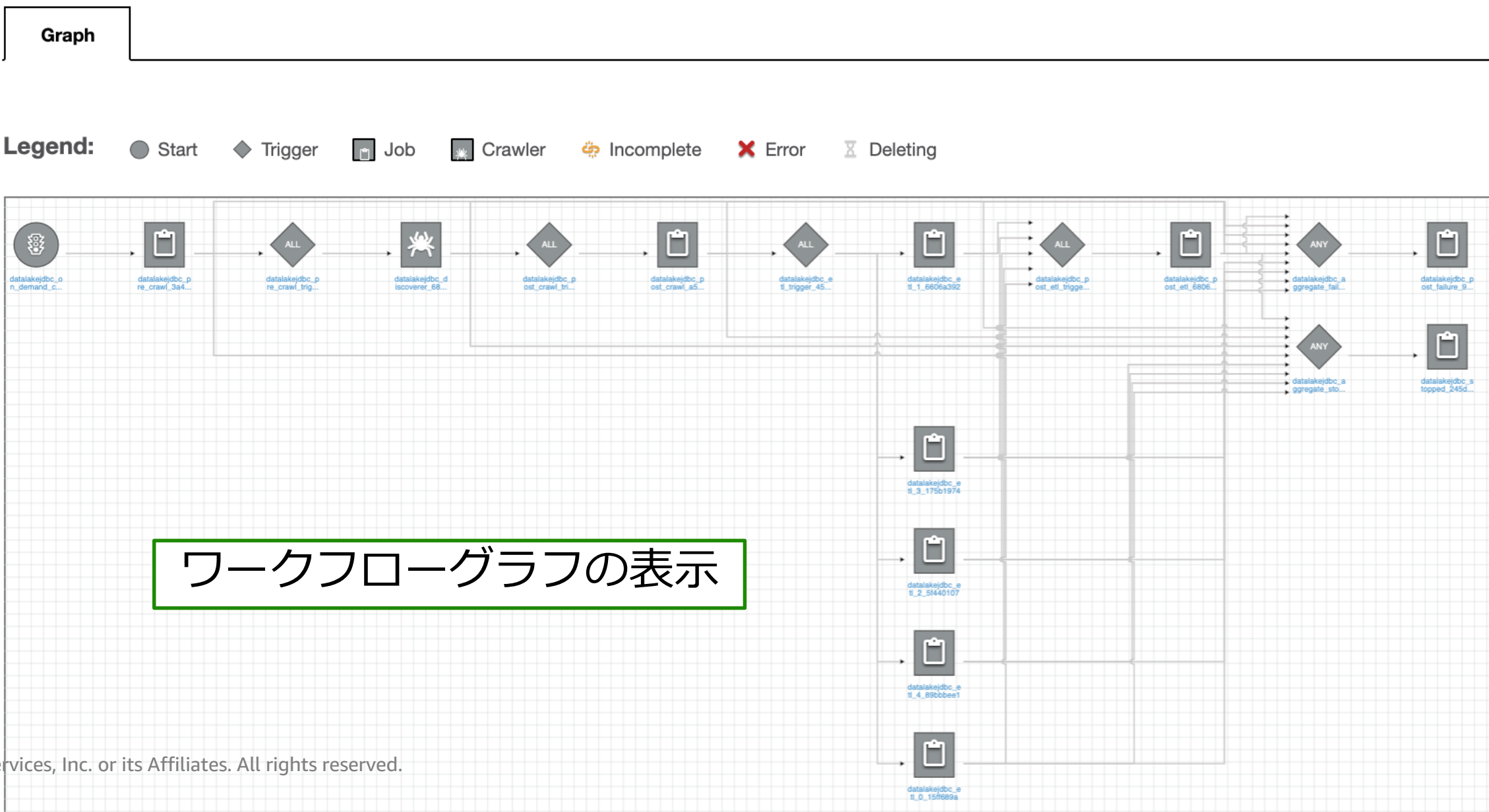
2019年8月24日(土) 13:26 UTC

wr_2b6e94ba0cfca4fbd84815ddc4032e3e71a1c646cd2be9e9f9fc5cc4984c677 [🔗](#)



ワークフロー：ブループリントから作成される処理フロー

Lake Formation のワークフローの実態は、Lake Formation のブループリントから生成される一連の Glue ジョブ、クローラー、トリガーの定義セットである Glue ワークフローです。ワークフローによって複雑な ETL ジョブアクティビティを隠蔽し、下記のようなワークフローグラフにより進捗を可視化することが可能です。



ユースケース : Database snapshot のブループリント

Database snapshot のブループリントを使うと、RDB のデータを分析に適したパーティションの最適化やフォーマット(parquetなど)変換を行い S3 に出だし Athena 等で分析が行えます。ブループリントに必要なパラメータを入力し、作成された Lake Formation のワークフローを実行することでデータセットやテーブルの作成は完了します。

Database snapshot のブループリントによりワークフロー作成

Import source
Configure the workflow source.

Database connection
Choose the connection to the data source. [Create a connection in AWS Glue](#)

datalake-tutorial

Source data path
Enter the path from which to ingest data. For JDBC databases with schema support, enter database/schema/table. Substitute the percent (%) wildcard for schema or table.

dblif/person

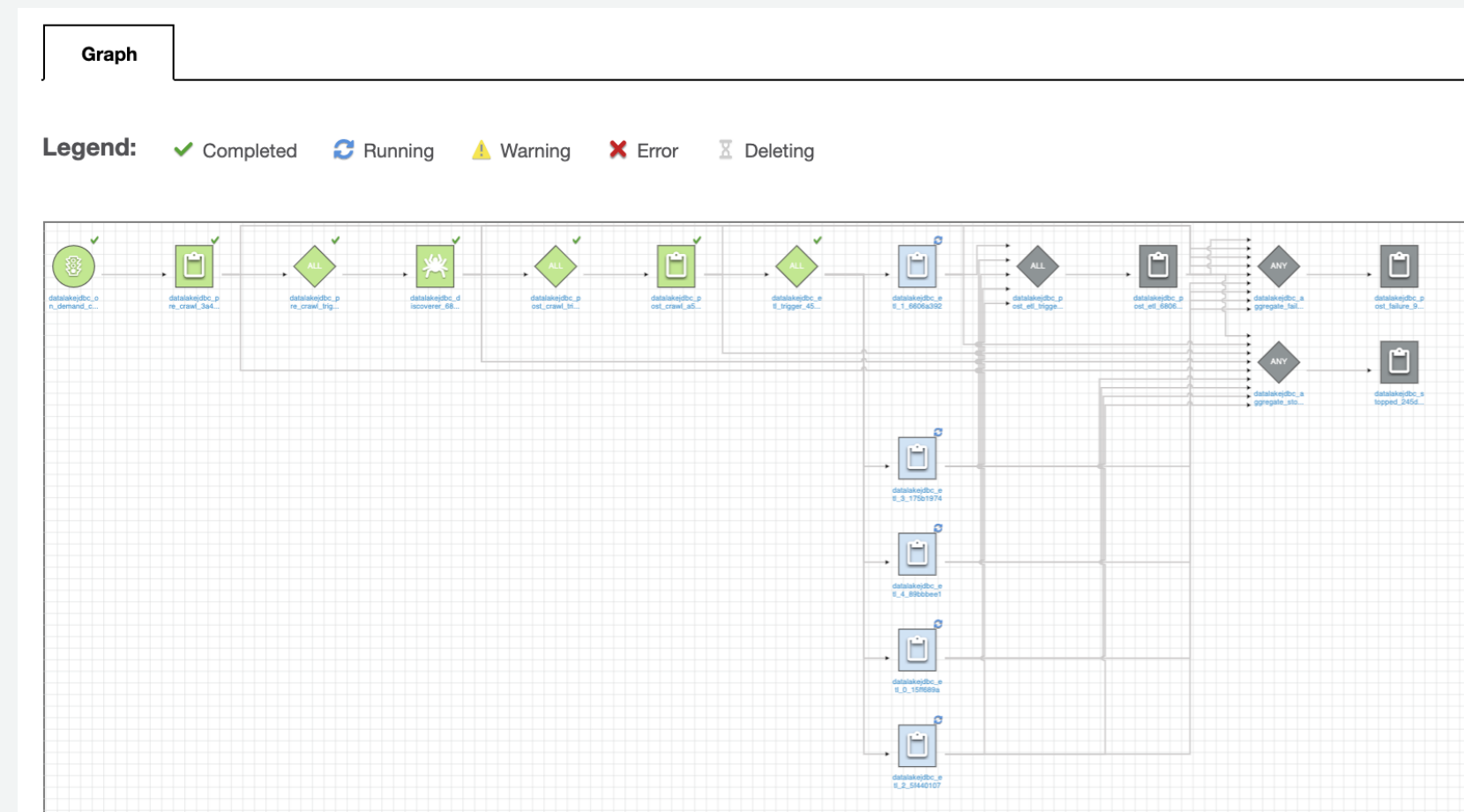
Exclude patterns - optional
Specify a glob pattern to limit the workflow. Tables that match this pattern will be excluded. [Learn more](#)

Exclude pattern

ワークフロー実行



ワークフローの進捗状況



ユースケース : Database snapshot のブループリント

利用者は見たいテーブルを簡単に探せます。テーブルにテーブルプロパティを キー:バリュー (例、env:research) で設定しておくことで、利用者が env:research で検索し閲覧権限のある該当のテーブルを見つけることができます。

Table properties

CreatedByJob	datalakejdbc_etl_3_175b1974	CreatedByJobRun	jr_7b76a21487f639f1262014a989b4e6bca1bdbca787aea6988849e127dc51b90						
LastTransformCompletedOn	2019-09-01 04:08:07.075201	LastUpdatedByJob	datalakejdbc_etl_3_175b1974						
LastUpdatedByJobRun	jr_2f2eb28235fe510980dacf37614941b7981951e00cc2f1519068b7342dd2c87d	SourceConnection	lf-3-connect						
SourceTableName	dblf_person	SourceType	JDBC	TableVersion	2	TransformTime	0:01:17.846746	classification	PARQUET
env	research								

AWS Lake Formation ✕

- Dashboard
- ▼ Data catalog
 - Databases
 - Tables**
 - Settings
- ▼ Register and ingest
 - Data lake locations

AWS Lake Formation Tables

Tables (1) 🔄 Actions ▼ Create table using a crawler ↗

keyword : env:research ✕

	Name ▼	Database ▼	Location ▼	Classific... ▼	Last updated
○	datalakejdbc_dblf_person	datalake_jdbc	s3://uehara-datalake-tutorial/...	PARQUET	2019年9月1日(日) 6:2

ユースケース : Database snapshot のブループrint

該当のテーブルにチェックを入れ、[Actions] から [View data] をクリックします。


The screenshot displays the AWS Lake Formation console interface. On the left, a navigation sidebar includes 'Dashboard', 'Data catalog', 'Databases', 'Tables', 'Settings', 'Register and ingest', 'Data lake locations', 'Blueprints', 'Crawlers', and 'Jobs'. The main content area is titled 'AWS Lake Formation Tables' and shows a search filter for 'env:research'. A table with one entry is visible, with the first row selected. The 'Actions' dropdown menu is open, and the 'View data' option is highlighted. The table entry details are as follows:

Name	Database	Location	File...	Last updated
<input checked="" type="checkbox"/> datalakejdbc_dblf_person	datalake_jdbc	s3://uehara-data...	JET	2019年9月1日(日) 6:29 UTC

ユースケース : Database snapshot のブループリント

Athena の画面に遷移し Athena で分析を行うことができます。

Athena クエリエディタ 保存したクエリ 履歴 AWS Glue Data Catalog ワークグループ : primary 設定 チュートリアル

データベース 

datalake_jdbc



テーブルとビューのフィルタリング...

▼ テーブル (2) [テーブルの作成](#)

- ▶ _temp_datalakejdbc_dbf_person
- ▶ datalakejdbc_dbf_person

▼ ビュー (0) [ビューの作成](#)

ビューをまだ作成していません。ビューを作成するには、クエリを実行し、[クエリからビューを作成]をクリックします。

新しいクエリ 1  新しいクエリ 2  +

1 SELECT * FROM "datalake_jdbc"."datalakejdbc_dbf_person" limit 10;

[クエリの実行](#) [名前を付けて保存](#) [作成](#) (実行時間: 3.46 秒, スキャンしたデータ: 0.86 KB) [クエリのフォ](#)

クエリの実行には Ctrl + Enter、オートコンプリートには Ctrl + Space を使用します

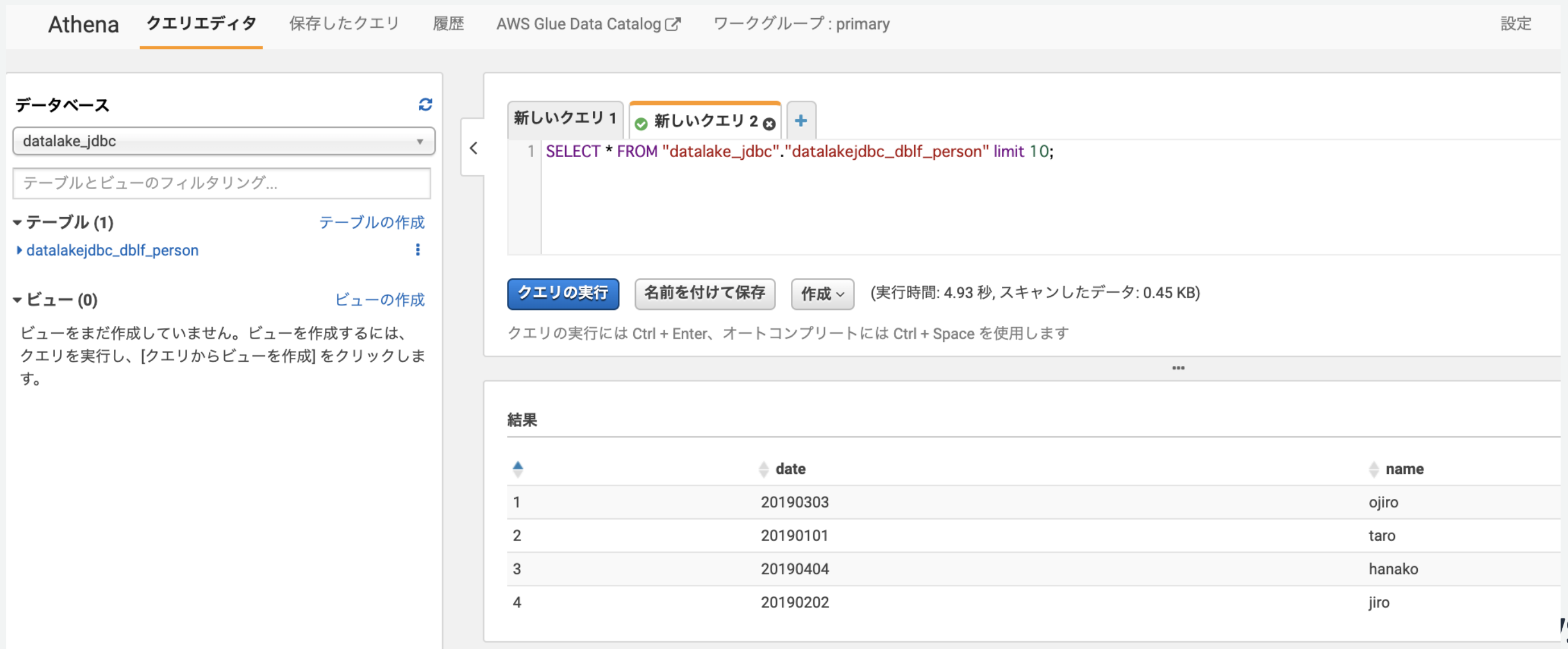
結果

	date	accountnumber	name	storeid
1	20190202	00001	jiro	102
2	20190101	00001	taro	101
3	20190404	00004	hanako	103
4	20190303	00002	ojiro	101

ユースケース : Database snapshot のブループリント

ユーザーごとに列レベルの制御を行い、アクセスできる列を絞ることで特定のユーザーに見せるデータを制御することも可能です。

以下の例は datalake_analyst というユーザーに date と name の列のみ select を許可しています。



The screenshot shows the Amazon Athena console interface. At the top, there are navigation tabs: Athena, クエリエディタ (Query Editor), 保存したクエリ (Saved Queries), 履歴 (History), AWS Glue Data Catalog, and ワークグループ : primary (Workgroup: primary). A 設定 (Settings) link is in the top right.

On the left side, under the 'データベース' (Database) section, 'datalake_jdbc' is selected. Below it, there's a search box for 'テーブルとビューのフィルタリング...' (Filtering tables and views...). Under 'テーブル (1)' (Tables (1)), 'datalakejdbc_dblf_person' is listed. Under 'ビュー (0)' (Views (0)), there's a message: 'ビューをまだ作成していません。ビューを作成するには、クエリを実行し、[クエリからビューを作成]をクリックします。' (You have not yet created any views. To create a view, run a query and click [Create view from query]).

The main area shows a query editor with two tabs: '新しいクエリ 1' (New Query 1) and '新しいクエリ 2' (New Query 2). The active query is: `1 SELECT * FROM "datalake_jdbc"."datalakejdbc_dblf_person" limit 10;`

Below the query editor, there are buttons: 'クエリの実行' (Execute Query), '名前を付けて保存' (Save with name), and '作成' (Create). The execution status shows: '(実行時間: 4.93 秒, スキャンしたデータ: 0.45 KB)'. A note below says: 'クエリの実行には Ctrl + Enter、オートコンプリートには Ctrl + Space を使用します' (Use Ctrl + Enter for query execution, Ctrl + Space for auto-completion).

The results section, titled '結果' (Results), shows a table with three columns: an index, 'date', and 'name'. The data is as follows:

	date	name
1	20190303	ojiro
2	20190101	taro
3	20190404	hanako
4	20190202	jiro

ユースケース : Database snapshot のブループrint

該当のテーブルの [Verify permissions] で datalake_analystist に与えられている権限を確認することができます。date と name の列のみ select が許可されていることがわかります。

The screenshot shows the AWS Glue console interface. In the 'Tables (2)' section, the table 'datalakejdbc_dblf_person' is selected. The 'Actions' menu is open, and 'Verify permissions' is highlighted. A modal window titled 'Verify permissions for datalakejdbc_dblf_person' is displayed, showing the IAM user 'datalake_analystist' and a table of permissions.

Resource type	Resource	Principal type	Principal	Permissions	Grantable
Column	Include: datalake_jdbc.data lakejdbc_db lf_person. [date, name]	IAM user	datalake_analystist	Select	-

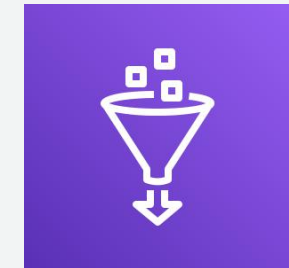
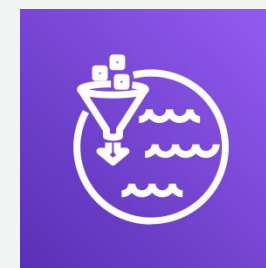
© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved.

落ち穂拾い

Lake Formation と Glue の関係

Glue は抽出、変換、ロード (ETL) を行うジョブとデータカタログを持つマネージドサービスです。

- Lake Formation は Glue の拡張機能であり、Glue に大きく依存しています。
- Lake Formation は Glue とデータカタログを共有しています。
- Lake Formation のジョブとクローラーは Glue のジョブとクローラーを呼び出します。
- Lake Formation のブループリントで生成されるワークフローは Glue のワークフローです。Lake Formation と Glue の両方でワークフローの表示と管理ができます。



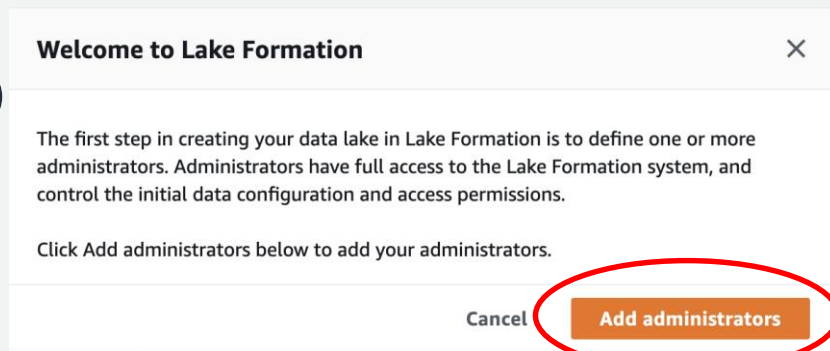
<https://www.slideshare.net/AmazonWebServicesJapan/20190806-aws-black-belt-online-seminar-aws-glue>

初めてデータレイク管理者を作成する

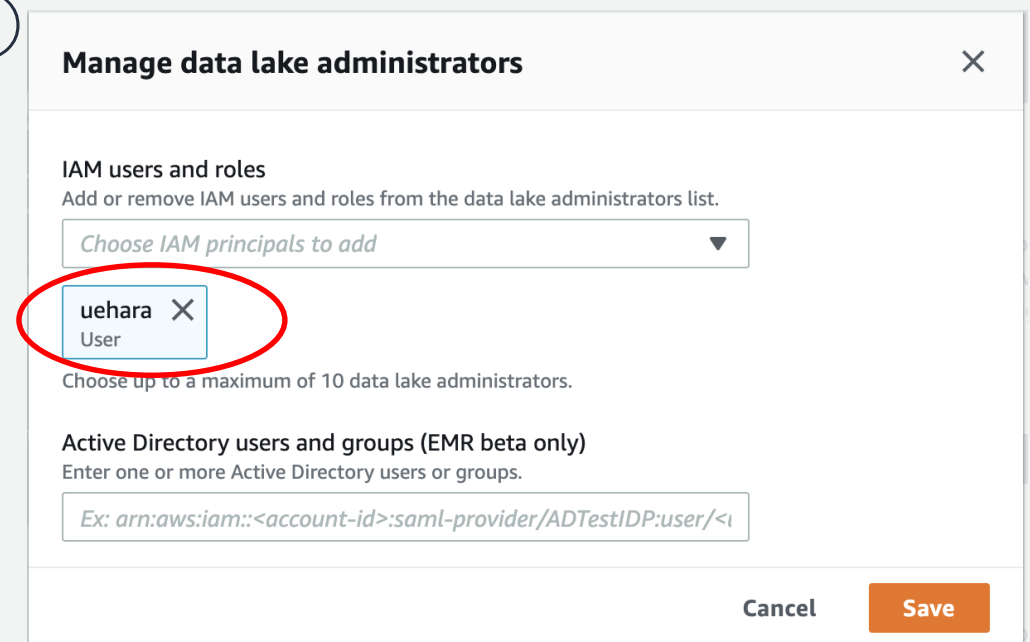
任意の IAM 管理ユーザーとしてサインインします。①または②を行った後③を実行します。

- ① Lake Formation に初めてアクセスするとウェルカムメッセージが表示されます。この画面で [Add administrator] を選択します。
- ② ナビゲーションペインで、[Admins and database creators] を選択します。次に、“Data lake administrators” で、[Grant] を選択します。
- ③ [Manage data lake administrators] ダイアログボックスの “IAM users and roles” で、データレイク管理者になるユーザーを選択し、[Save] を選択します。

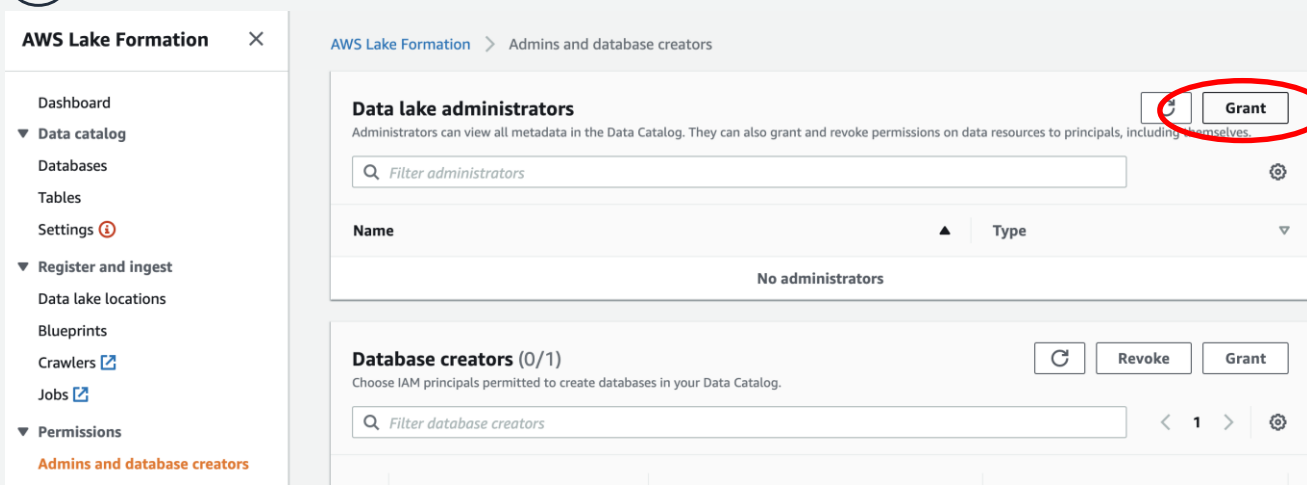
①



③



②



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

料金

例として、ブループリント“Database snapshot”をデフォルト設定で RDS から200MBのテーブルを S3 にロードし、Athena で100回フルスキャンクエリを利用した場合

- Glue クローラー1つを数分実行
最小時間の10分の料金： $\$0.44 / 6 = \0.0734
- Glue Spark ジョブ 5つを数分実行
最小時間の10分の料金* 5 ジョブ： $(\$0.44 / 6) * 5 = \0.367
- Glue Python Shell ジョブ5つを数秒実行
最小時間の1分の料金* 5 ジョブ： $(\$0.44 / 60) * 5 = \0.0367
- S3 に200MBのファイルを保存
1GBあたり\$0.025なので200MB換算： $\$0.005$
- Athena で100回フルスキャンクエリを実行
1 TBで\$5なので、0.0002TB(200MB)を掛けて回数100を掛ける： $(\$5 * 0.0002) * 100 = \0.1

✓ **合計で\$ 0.5821**

※本日時点の東京リージョンの料金の例です。正確には公式の料金のページを確認ください

※実際の処理時間はデータの内容により変わります

※Glue データカタログ、S3 のリクエスト料金は少額なので割愛しています

※Lake Formation によって作成されるリソースの数や内容は今後変更される可能性があります

<https://aws.amazon.com/jp/athena/pricing/>

<https://aws.amazon.com/jp/s3/pricing/>

<https://aws.amazon.com/jp/glue/pricing/>

まとめ

AWS Lake Formation まとめ

- ブループリントと呼ばれる汎用的なデータ取込み処理のテンプレート化
 - Glue のワークフロー、クローラー、ジョブと連携
- データのアクセス権限管理をシンプルに直感的に設定可能
 - データカタログのデータベース、テーブル、列に対して、Grant/Revoke 設定
 - メタデータ、S3 のデータレイクストアへのアクセス制御の一元管理
 - Athena、Redshift Spectrum、EMR(今後予定)からのアクセスを容易に
 - 既存の Glue データカタログからは移行が必要
- Lake Formation 自体の費用は無償
 - 配下で使用するサービスの費用のみ

参考資料

AWS Lake Formation ホームページ

<https://aws.amazon.com/jp/lake-formation/>

AWS Lake Formation ドキュメンテーション (開発者ガイド)

https://docs.aws.amazon.com/ja_jp/lake-formation/

AWS Lake Formation FAQ

<https://aws.amazon.com/jp/lake-formation/faqs/>

データレイクとは

<https://aws.amazon.com/jp/big-data/datalakes-and-analytics/what-is-a-data-lake/>

AWS Lake Formation API

<https://docs.aws.amazon.com/lake-formation/latest/dg/aws-lake-formation-api.html>

CLI Command Reference

<https://docs.aws.amazon.com/cli/latest/reference/lakeformation/index.html#cli-aws-lakeformation>

Cloud Formation Lake Formation Resource Type Reference

https://docs.aws.amazon.com/AWSCloudFormation/latest/UserGuide/AWS_LakeFormation.html

参考資料(公式ブログ)

AWS Lake Formation の開始方法

<https://aws.amazon.com/jp/blogs/news/getting-started-with-aws-lake-formation/>

AWS Lake Formation でデータレイクを構築、保護、管理

<https://aws.amazon.com/jp/blogs/news/building-securing-and-managing-data-lakes-with-aws-lake-formation/>

AWS Lake Formation でメタデータを見つける: パート 1

<https://aws.amazon.com/jp/blogs/news/discovering-metadata-with-aws-lake-formation-part-1/>

AWS Lake Formation でメタデータを見つける: パート 2

<https://aws.amazon.com/jp/blogs/news/discover-metadata-with-aws-lake-formation-part-2/>

AWS Lake Formation FindMatches を使用してデータセットの統合および重複の削除を実施

<https://aws.amazon.com/jp/blogs/news/integrate-and-deduplicate-datasets-using-aws-lake-formation-findmatches/>

Q&A

お答えできなかったご質問については

AWS Japan Blog 「<https://aws.amazon.com/jp/blogs/news/>」にて
資料公開と併せて、後日掲載します。

AWS の日本語資料の場所「AWS 資料」で検索

AWS クラウドサービス活用資料集トップ

アマゾン ウェブ サービス (AWS) は安全なクラウドサービスプラットフォームで、ビジネスのスケールと成長をサポートする処理能力、データベースストレージ、およびその他多種多様な機能を提供します。お客様は必要なサービスを選択し、必要な分だけご利用いただけます。それらを活用するために役立つ日本語資料、動画コンテンツを多数ご提供しております。(本サイトは主に、AWS Webinar で使用した資料およびオンデマンドセミナー情報を掲載しています。)

[AWS Webinar お申込 »](#)

[AWS 初心者向け »](#)

[サービス別資料 »](#)

<https://amzn.to/JPArchive>

ご視聴ありがとうございました