

# AWS Lake Formation 入門ワークショップ

2021年11月

アマゾン ウェブ サービス ジャパン 合同会社  
シニアソリューションアーキテクト  
下佐粉 昭 (Akira Shimosako)

 @simosako

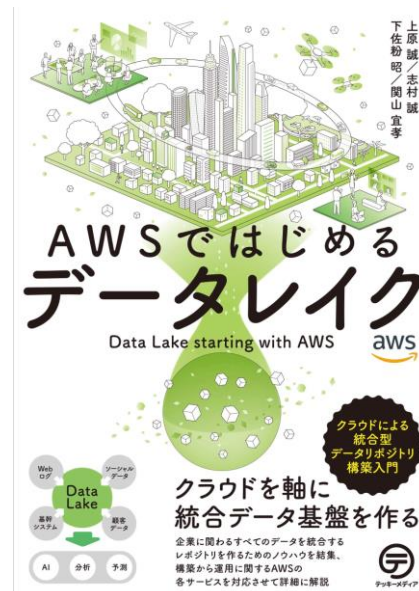
# 自己紹介

下佐粉 昭 (しもさこ あきら)  
アマゾン ウェブ サービス ジャパン  
シニアソリューションアーキテクト (アナリティクス)

 @simosako

「AWSではじめるデータレイク」  
<https://techiemedia.co.jp/>

週刊AWS :  
毎週AWSのアップデートをサマリしてお届け



# 本ワークショップの内容

- AWS Lake Formationの基本的な概念や操作をAWS環境で体験しながら学習するワークショップです
- ワークショップ実施には**所定のAWS利用料金が発生します**
- 内容は以下のAWS Lake Formation Workshop の内容を一部変更し、日本語の説明を付けたものです
  - <https://lakeformation.workshop.aws/>
- この時間内では上記URLの内容全ては実施できませんが、本ワークショップ完了後に、続きを実施いただく事は可能です

# ワークショップを実施するために必要になる環境

- AWSアカウント
  - AdministratorAccessポリシー等の強い権限をもつAWS Adminユーザ（ロール）でのログインが可能であること
- PCとブラウザ（Firefox、Chrome、Edge等）
- インターネット接続
- ワークショップコンテンツ
  - ワークショップ説明スライド（本資料）
  - lf-workshop-mod.cf.yaml（CloudFormationテンプレート）
  - nyctaxi-schema.txt（nyctaxiの表定義ファイル）

# 補足：利用AWSリージョンについて

本資料では利用するAWSリージョンとして  
バージニア北部(us-east-1)を想定して記載しています

ワークショップは、Lake Formationが利用できる環境であれば他リージョンでも実行可能になっていますので、必要に応じて別リージョンを利用いただく事は可能です

- ・ バージニア北部以外ではオレゴン(us-west-2)での実行を確認済み

バージニア北部以外を使用する場合は、本資料の  
バージニア北部と書かれた部分を適宜読み替えてください

# ワークシヨツプ環境の準備 (AWS Admin権限のあるユーザで実施)

# (必要な方のみ)

## EC2キーペアが無い場合は作成する

ワークショップ環境を構築するためにEC2のキーペアが必要です（秘密鍵を保持している必要はありません）。使用リージョンに**EC2キーペアが存在しない場合は以下にしたがってキーペアを作成**してください。

AWS側で用意したAWS環境の場合、キーペアは作成済なのでここで作成する必要はありません。

AWS管理コンソールにログインし、利用リージョンが「**バージニア北部**」である事を確認してください

- EC2コンソールを開く
- 画面左の「キーペア」を選択
- 「キーペアを作成」を押す
- 名前を付ける（半角英数字）
- プライベートキーファイル形式は任意で選択
- 画面下部の「キーペアを作成」を押す

### キーペアを作成 情報

#### キーペア

プライベートキーとパブリックキーの組合せで構成されるキーペアは、セキュリティ認証情報としてインスタンス接続時の ID 証明に使用されます。

名前

lfworkshop

名前には最大 255 文字の ASCII 文字を使用できます。先頭または末尾のスペースを含めることはできません。

キーペアのタイプ 情報

☒ RSA

☐ ED25519

プライベートキーファイル形式

☒ .pem

OpenSSH で使用する場合

☐ .ppk

PuTTY で使用する場合

タグ(オプション)

リソースにタグが関連付けられていません。

タグを追加

さらに 50 のタグを追加できます

キャンセル

キーペアを作成

# CloudFormationによる環境構築

利用リージョンが「**バージニア北部**」である事を念のために確認してください

- AWS管理コンソールで CloudFormation を開く
- 「スタックの作成」をクリック
- テンプレートソースとして「テンプレートファイルのアップロード」を選択
- 配布された「**lf-workshop-mod.cf.yaml**」をアップロードして、「次へ」

## スタックの作成

### 前提条件 - テンプレートの準備

#### テンプレートの準備

各スタックはテンプレートに基づきます。テンプレートとは、スタックに含む AWS リソースに関する設定情報を含む JSON または YAML ファイルです。

☒ テンプレートの準備完了

☐ サンプルテンプレートを使用

☐ デザイナーでテンプレートを作成

### テンプレートの指定

テンプレートは、スタックのリソースおよびプロパティを表す JSON または YAML ファイルです。

#### テンプレートソース

テンプレートを選択すると、保存先となる Amazon S3 URL が生成されます。

☐ Amazon S3 URL

☒ テンプレートファイルのアップロード


#### テンプレートファイルのアップロード

ファイルの選択  lf-workshop-mod.cf.yaml

JSON または YAML 形式のファイル

S3 URL: https://s3-external-1.amazonaws.com/cf-templates-eia84vnmvixt-us-east-1/

kshop-mod.cf.yaml

デザイナーで表示 

キャンセル

次へ



# CloudFormationによる環境構築

- 最上部のスタック名には任意の名前を入力 (lfworkshop 等)
- EC2 Key Pairでは、存在するキーペアから1つ選択
  - ワークショップではEC2にログインしませんのでどのキーペアでも問題ありません
  - AWSが主催しているイベントの場合は**ee-default-keypair**が存在しますので、そちらを選択してください
- 他はデフォルトのままで「次へ」
- スタックオプションの設定ページでは特に変更せず「次へ」

**パラメータ**

パラメータは、テンプレートで定義されます。また、パラメータを使用すると、スタックを作成または更新する際にカスタム値を入力できます。

**Database Configuration**

**Database Name**  
Name of the database that will be created.

tpc

**Master Username**  
Master username for TPC database.

tpcadmin

**Master User Password**  
Master password for TPC database.

**User Configuration**

**Test User Password**  
Password for all test users.

**Misc Configuration**

**EC2 Key Pair**  
Amazon EC2 Key Pair

ee-default-keypair

**Latest AMI Id**  
Image ID for the EC2 helper instance. DO NOT change this.

/aws/service/ami-amazon-linux-latest/amzn2-ami-hvm-x86\_64-gp2



# CloudFormationによる環境構築

レビューのページでは内容を確認し、最下段の「AWS CloudFormation によって IAM リソースがカスタム名で作成される場合があることを承認します。」に**必ずチェックを入れて**「スタックの作成」

スタックが作成されるのに約6分、その後EC2やRDSの準備が整うまで7分ほど、**計13分程度**かかります

**i The following resource(s) require capabilities: [AWS::IAM::User, AWS::IAM::InstanceProfile]**

このテンプレートには、Identity and Access Management (IAM) リソースが含まれています。これらのリソースを個別に作成し、それぞれに最小限必要な権限を付与する必要があります。さらに、カスタム名が付けられているか確認してください。カスタム名が、ご利用の AWS アカウント内で一意のものであることを確認してください。 [詳細はこちら](#)

☒ **AWS CloudFormation によって IAM リソースがカスタム名で作成される場合があることを承認します。**

キャンセル 戻る 変更セットの作成 **スタックの作成**

重要！

CloudFormationでは主に以下のリソースが準備されます

- [illegible]

# AWS Adminユーザと CloudFormationで作成されるIAMユーザー

※IAMユーザーのパスワードは全て同じ

※パスワードはCloudFormationの「出力」タブで確認可能

	役割	補足
AWS Admin (ユーザもしくはロール)	Lake Formation 管理者が利用可能になるまでの環境準備	<b>ワークショップ参加者側で準備する</b> AWS側で環境を用意している場合は“TeamRole” ロール
If-admin (IAM user)	Data lake (Lake Formation) 管理者	Lake Formationでの権限設定はこのIAMユーザで行う
If-developer (IAM user)	管理者に許可された範囲でLake Formationを利用する	
If-campaign-manager (IAM user)	管理者に許可された範囲でLake Formationを利用する	
If-business-analyst (IAM user)	管理者に許可された範囲でLake Formationを利用する	※本セッションでは利用しない
LF-GlueServiceRole (IAM role)	ワークフローを実行するためのロール	Lake Formationのワークフローは、Glue を利用して作成される

# CloudFormationの出力タブを確認

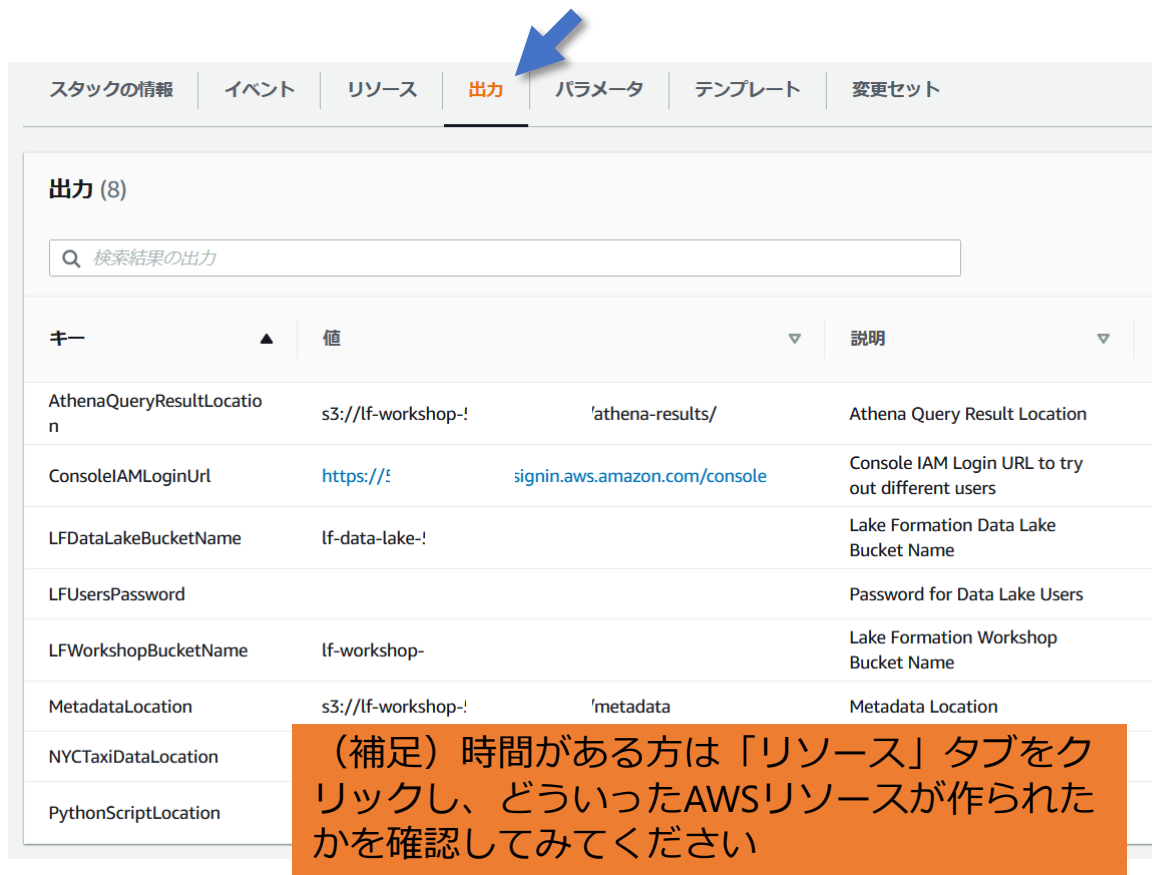
CloudFormationの「出力」を選択し、内容を確認します

AthenaQueryResultLocationはAthena利用時に設定するS3 URLです。**値に表示されたURLをメモしておいてください**

ConsoleIAMLoginURLはワークショップの途中でIAMユーザに切り替える際のログインURLです。**値に表示されたURLをメモしておいてください**

LFDataLakeBucketNameは、データレイクのデータ置き場となるバケットです。**lf-data-lake-...**から始まるバケット名です

LFUsersPasswordは、IAMユーザ用のパスワードです。**値に表示されたパスワードをメモしておいてください**



スタックの情報 | イベント | リソース | **出力** | パラメータ | テンプレート | 変更セット

出力 (8)

検索結果の出力

キー	値	説明
AthenaQueryResultLocation	s3://lf-workshop-! 'athena-results/'	Athena Query Result Location
ConsoleIAMLoginURL	https://! signin.aws.amazon.com/console	Console IAM Login URL to try out different users
LFDataLakeBucketName	lf-data-lake-!	Lake Formation Data Lake Bucket Name
LFUsersPassword		Password for Data Lake Users
LFWorkshopBucketName	lf-workshop-	Lake Formation Workshop Bucket Name
MetadataLocation	s3://lf-workshop-! 'metadata	Metadata Location
NYCTaxiDataLocation		
PythonScriptLocation		

(補足) 時間がある方は「リソース」タブをクリックし、こういったAWSリソースが作られたかを確認してみてください

# Lake Formation環境のセットアップ (AWS Adminでの作業)

# Lake Formation管理者の設定

Lake Formationの初期セットアップを行います。  
最初にLake Formation管理者を設定（登録）します

※Lake Formationコンソールにアクセスした際に右図Welcomeダイアログが出ない場合は、次ページを参照してください

本ワークショップではAWS Adminユーザと、If-admin (IAM user)をLake Formation管理者として登録します

Lake Formationコンソールに移動します  
右のダイアログが出るので

- Add myselfをチェック
- Add other AWS users or rolesをチェックし、If-adminを管理者として登録

「Get started」をクリック

## Welcome to Lake Formation

The first step in creating your data lake in Lake Formation is defining one or more administrators. Administrators have full access to the Lake Formation console, and control the initial data configuration and access permissions.

### Choose the initial administrative users and roles

You may add yourself and/or other principals.

☒ Add myself

AWS account:

☒ Add other AWS users or roles

Select additional IAM users and roles to be data lake administrators.

Choose IAM principals to add

If-admin X  
User

Choose up to a maximum of 10 data lake administrators.

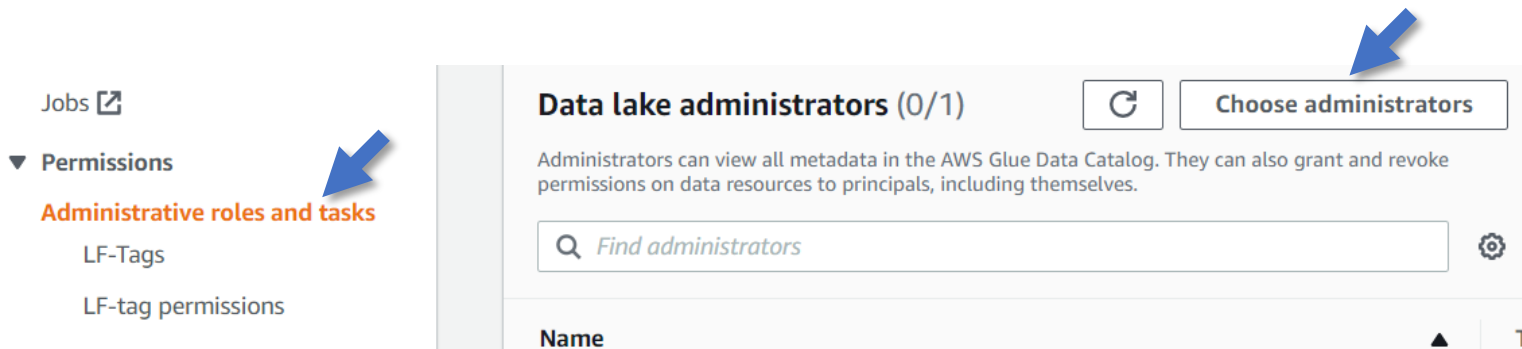
Cancel

Get started

# 補足：Lake Formationコンソールにアクセスした際に、ウェルカムダイアログが出ない場合

Lake Formationコンソールにアクセスした際に  
前ページのようなウェルカムダイアログが出ない場合、  
その環境はすでにData lake (Lake Formation)管理者が登録済である可能性があります

その場合はLake Formationの”Administrative roles and tasks”画面で、  
**lf-admin**を**Data lake administrators**に追加してください





# Lake Formationの設定

デフォルトではIAMベースで動いているアプリケーションに影響が出ないように、IAMアクセスと互換性があるように設定されています

本ワークショップではLake Formationベースの権限管理を行うため、上記を無効にします

- 管理画面の左側でSettingsを選択
- Use only IAM access control for new databasesからチェックを外す
- Use only IAM access control for new tables in new databasesからチェックを外す
- Saveを選択

補足：この時点で画面下部に「Your account is not a member of an organization.」や「You don't have permissions to access this resource.」と出ても本ワークショップでは問題ありません

AWS Lake Formation > Data catalog settings

## Data catalog settings

### Default permissions for newly created databases and tables

These settings maintain existing AWS Glue Data Catalog behavior. You can still set individual permissions on databases and tables, which will take effect when you revoke the Super permission from IAMAllowedPrincipals. See [Changing Default Settings for Your Data Lake](#).

☐ Use only IAM access control for new databases

☐ Use only IAM access control for new tables in new databases

### Default permissions for AWS CloudTrail

These settings specify the information being shown in AWS CloudTrail.

**Resource owners**  
Enter resource owners you wish to share your CloudTrail access details with.

Enter one or more AWS account IDs. Press Enter after each ID.

**❗** You don't have IAM permissions to make cross-account grants. The required permissions are in the AWS managed policy AWSLakeFormationCrossAccountManager.

**❌** You don't have permissions to access this resource.

Cancel Save

# Lake Formationの設定

続いて左側からAdministrative roles and tasksを選択し、最下部の”IAMAllowedPrincipals”を選択し、Revokeを選択します  
ダイアログが出るので、そのまま”Revoke”を押します

## ▼ Permissions

### Administrative roles and tasks



LF-Tags

LF-tag permissions

Data lake permissions



Data locations

External data filtering

**Data lake administrators (0/2)**   Choose administrators

Administrators can view all metadata in the AWS Glue Data Catalog. They can also grant and revoke permissions on data resources to principals, including themselves.

Name	Type
lf-admin	IAM user
TeamRole	IAM role

**Database creators (0/1)**   Revoke Grant

Choose IAM principals permitted to create databases in your AWS Glue Data Catalog.


Principal	Principal type	Permissions	Grantable
<input checked="" type="radio"/> IAMAllowedPrincipals	Group	Create database	-

## Revoke permissions

Revoke access permissions to specific users and roles.

### IAM users and roles

Add one or more IAM users or roles.

**IAMAllowedPrincipals**   
Group

### SAML and Amazon QuickSight users and groups

Enter a SAML user or group ARN or Amazon QuickSight ARN. Press Enter to add additional ARNs.

### Catalog permissions

Choose the access permissions to revoke. Access will be blocked even if IAM permissions are in place.

☒ Create database

### Grantable permissions

Choose the permissions that may not be granted to others.

☐ Create database

Cancel

Revoke 

# データレイクロケーションの登録

データレイクのデータを蓄積する「データレイクロケーション」を設定します

- Data lake locationsを選択  
(Data locationsと混同しないよう注意)
- Register locationを押す
- Browseをクリックし、CloudFormationの出力の「LFDataLakeBucketName」に表示されていた”lf-data-lake-...”バケットを選択
- IAM Roleはデフォルトのまま
- Register locationを選択

AWS Lake Formation > Data lake locations > Register location

## 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://lf-data-lake-

Browse

#### 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.

Review location permissions

#### 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



Do not select the service linked role if you plan to use EMR.

Cancel

Register location

# 補足：データレイクロケーション設定を AWS Adminで実施した理由

今回のワークショップではデータレイク管理者のlf-adminにAWSサービスへのアクセス許可を委任するアクション(iam:PutRolePolicy)が許可されていないため、データレイクロケーションの登録はAWS Adminで実施しています

# Lake Formation管理者(If-admin) によるデータレイク権限設定

# ここからの作業はlf-adminユーザで実施します

ブラウザにはすでにAWS Adminでログインしているため、プライベートウィンドウを利用するか、他のブラウザを利用してlf-adminとしてAWS管理コンソールにログインします

ログイン用のURLはCloudFormationの出力の"ConsoleIAMLoginUrl"で確認できます

- IDはlf-admin
- パスワードはCloudFormationの出力の"LFUsersPassword"に記載

例"ConsoleIAMLoginUrl"のURLを右クリックし、"リンクを新しいプライベートウィンドウで開く"をクリック

キー	値	説明	エクスポート名
AthenaQueryResultLocation	s3://lf-workshop-3/athena-results/	Athena Query Result Location	-
ConsoleIAMLoginUrl	<a href="https://0231011...signin.aws.amazon.com/console">https://0231011...signin.aws.amazon.com/console</a>	Console IAM Login URL to try	-
LFDataLakeBucketName	lf-data-lake-		-
LFUsersPassword			-
LFWorkshopBucketName	lf-workshop-		-
MetadataLocation	s3://lf-workshop-metadata		-
NYCTaxiDataLocation	s3://lf-workshop-glue/nyctaxi		-
PythonScriptLocation	s3://lf-workshop-glue/scripts/nyctaxi-to-json.py		-

# If-adminユーザでのログインを確認 リージョンに注意！



画面右上のユーザ名がIf-admin @ ...になっていることを確認

同様にリージョンを必ず確認し、「バージニア北部」（ワークショップ実施リージョン）を選択してください

（別ユーザにログインしなおすタイミングで初期設定リージョンが変わることがあるためです）

# If-adminユーザで実施する内容

ここではlf-admin (Lake Formation管理者)として以下の設定を行っていきます

- Lake Formation上にtpcデータベースを作成
- Blueprintを使い、RDBからデータレイクにデータを取り込むワークフローの作成 (RDB to S3)
- (ワークショップ時間短縮のため) 事前に用意されたS3上のnyctaxi表をLake Formationのデータベース内に表として手動登録
- Lake Formation利用ユーザに対しての権限設定
  - lf-developerには、**ユーザ名で指定して**表・列単位のアクセスを許可
  - lf-campaign-managerには、**LFタグ**を使って表・列単位のアクセスを許可



# データベースの作成

Lake Formation上にデータベースを作成します（データベースという管理単位であり、RDBMSが作成されるわけではありません）

- Lake Formationの画面を開く
- 左側のDatabasesをクリック
- Create databaseボタンを押す
- Nameにtpcと入力
- Browseでデータレイクロケーションと同じURL(lf-data-lake...)を設定
- Create databaseをクリック

※AWSが提供する環境の場合、ここで“Unknown error”が表示される事がありますが、そのまま進めても問題ありません

AWS Lake Formation > Databases > Create database

## Create database

**Database details**  
Create a database in the AWS Glue Data Catalog.

☒ **Database**  
Create a database in my account.

☐ **Resource link**  
Create a resource link to a shared database.

**Name**  
tpc

**Location - optional**  
Choose an Amazon S3 path for this database, which eliminates the need to grant data location permissions on catalog table partitions that are this location's children

s3://lf-data-lake Browse

**Description - optional**  
Enter a description

Descriptions can be up to 2048 characters long.

**Default permissions for newly created tables**  
This setting maintains existing AWS Glue Data Catalog behavior. You can still set individual permissions, which will take effect when you revoke the Super permission from IAMAllowedPrincipals. See [Changing Default Settings for Your Data Lake](#).

☐ Use only IAM access control for new tables in this database

Cancel Create database

# 補足 : data locationの権限について

今回のワークショップでは説明を省いていますが、Lake Formationでは、**data lake location**と**data location**の権限という概念があります

- **data lake location**はLake Formationの管理下におかれるロケーション(S3 URL)です
- **data location**の権限は、実際のデータが置かれたロケーション(S3 URL)に対し、そのデータをDB上に登録(CREATE TABLE)するために必要となる権限です
- CREATE DB時にLocationを指定すると、上記**data location**の権限設定が不要になります（今回のハンズオンでは手順を省略するためにこの方法をとっています）

詳細は以下のURLをご覧ください

<https://docs.aws.amazon.com/lake-formation/latest/dg/access-control-underlying-data.html>

## Dashboard

### ▼ Data catalog

Databases

Tables

Settings

### ▼ Register and ingest

Data lake locations

Blueprints

Crawlers 

Jobs 

### ▼ Permissions

Administrative roles and tasks

LF-Tags

LF-tag permissions

Data lake permissions

Data locations

External data filtering

# Blueprintによるワークフローの作成

BlueprintでRDBデータをデータレイクに取り込むワークフロー（ETLジョブ）を作成します

- 画面左側よりBlueprintを選択
- Use blueprintを押す

**AWS Lake Formation** ×

**Data catalog**

- Dashboard
- Data catalog**
- Databases
- Tables
- Settings

**Register and ingest**

- Data lake locations
- Blueprints**
- Crawlers
- Jobs

**Permissions**

**AWS Lake Formation** > Blueprints

**▼ Blueprint overview**  
Blueprints enable data ingestion from common sources using automated workflows.

**Database blueprints**  
Ingest data from MySQL, PostgreSQL, Oracle, and SQL server databases to your data lake, either as bulk load snapshot, or incrementally load new data over time.

**Log file blueprints**  
Ingest data from popular log file formats from AWS CloudTrail, Classic Load Balancer, and Application Load Balancer logs

**Use blueprint**

**Workflows**

**Use blueprint**

# Blueprintによるワークフローの作成

- Blueprint typeではDatabase snapshotを選択
- Import sourceにCloudFormationで事前作成したRDSへの接続を指定
  - Database connection: TPCGlueConnectorを選択
  - Source data path: tpc/ と入力 （※tpcというデータベースの表全てを選択）
- Exclude patternsはデフォルトのまま （次ページへ続く）

**Import source**  
Configure the workflow source.

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

TPCGlueConnector

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

tpc/

# Blueprintによるワークフローの作成

Import target :

- Target database : tpcを選択
  - Target storage location : (tpcを選択すると自動入力)
  - Data format: Parquetを選択
- (次ページへ続く)

### Import target

Configure the target of the workflow.

**Target database**  
Choose a database in the AWS Glue Data Catalog. [Create database](#)

tpc

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

s3://lf-data-lake-023101160108

Browse

**Data format**  
Choose the output data format.

Parquet

# Blueprintによるワークフローの作成

Import frequency :  
(変更しない)

Import Options :

- Workflow nameにはtpc-workflowと入力
- IAM roleにはLF-GlueServiceRoleを選択
- Table prefixにはdlと入力
- その他はデフォルトままで、“Create”をクリック

### Import options

Configure the workflow.

Workflow name

Name may contain letters (A-Z), numbers (0-9), hyphens (-), or underscores (\_), and must be less than 256 characters long.

IAM role

LF-GlueServiceRole ▼

Table prefix

The table prefix that is used for catalog tables that are created.

Table prefixes may contain lower case letters (a-z), numbers (0-9), hyphens (-), or underscores (\_).

Maximum capacity - *optional*

Sets the number of data processing units (DPUs) that can be allocated when this job runs. A DPU is a relative measure of processing power that consists of 4 vCPUs of compute capacity and 16 GB of memory.

Concurrency - *optional*

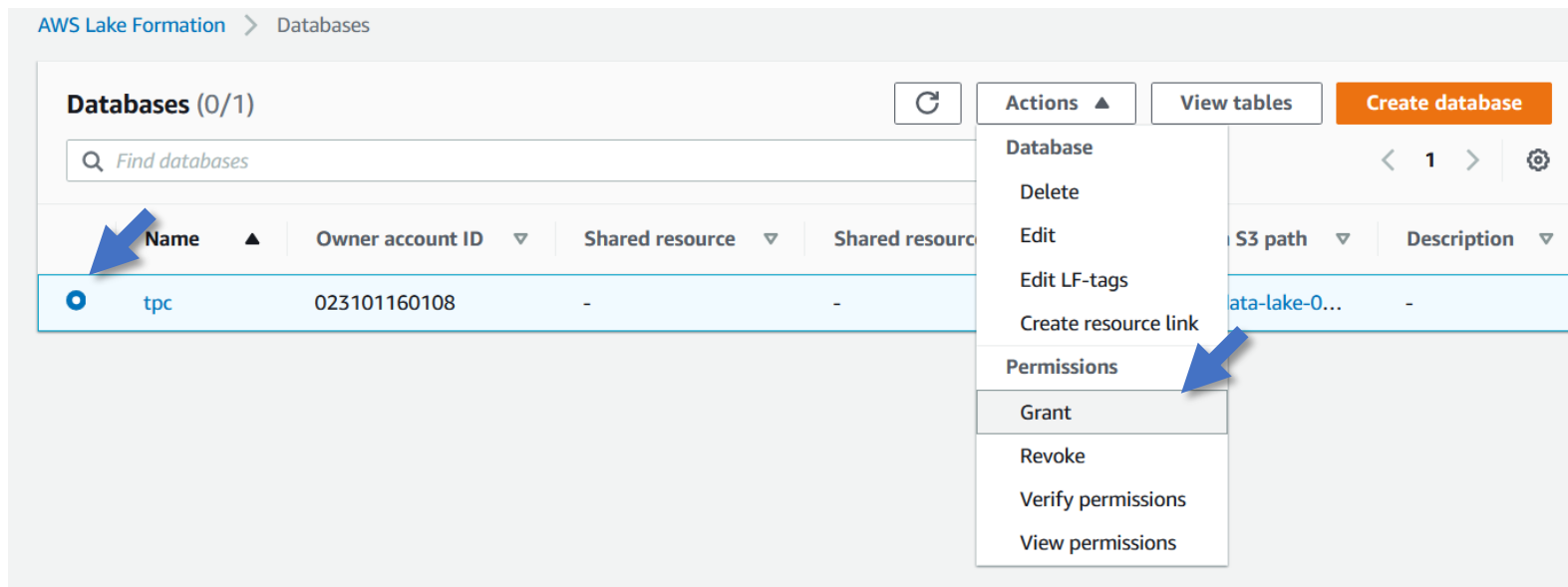
Sets the maximum number of concurrent runs that are allowed for this job. An error is returned when this threshold is reached. The default is 5.

Cancel Create

# LF-GlueServiceRoleにTPC DBアクセスを許可

作成したワークフローを実行する前に、ワークフローを実行するLF-GlueServiceRoleがtpcデータベースに書き込めるよう、権限を付与します

- 画面左のDatabasesを選択
- tpcにチェックをいれて、Actions -> Grantを選択



# LF-GlueServiceRoleにTPC DBアクセスを許可

lf-admin

## Principals:

- IAM users and rolesを選択
- 一覧からLF-GlueServiceRoleを選択

## LF-Tags or catalog resources

- Named data catalog resourcesで、tpcが選択済になっているはずなのでそのまま使用

(次ページに続く)

### Grant data permissions

#### Principals



**IAM users and roles**

Users or roles from this AWS account.



**SAML users and groups**

SAML users and group or QuickSight ARNs.



**External accounts**

AWS accounts or AWS organizations outside of this account.

#### IAM users and roles

Add one or more IAM users or roles.

Choose IAM principals to add

LF-GlueServiceRole X  
Role

#### LF-Tags or catalog resources



**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**

Manager permissions for specific databases or tables, in addition to fine-grained data access.

#### Databases

Select one or more databases.

Choose databases

tpc X  
555085981125

#### Tables - optional

Select one or more tables.

Choose tables

Load more

Load more



# LF-GlueServiceRoleにTPC DBアクセスを許可

If-admin

## Database permissions

- Database permissionsの中の”Create table”にチェックをいれて、”Grant”を押す

※下図のようにLF-GlueServiceRoleにパーミッションが付与された事が分かる

**Database permissions**

Database permissions  
Choose specific permissions to grant.

☒ Create table ☐ Alter ☐ Drop  
☐ Describe

**Grantable permissions**  
Choose the permission that may be granted to others.

☐ Create table ☐ Alter ☐ Drop  
☐ Describe

☐ Super  
This permission is the union of all the individual permissions to the left, and supersedes them.

☐ Super  
This permission allows the principal to grant any of the permissions to the left, and supersedes those grantable permissions.

Cancel Grant

**Data permissions (2)**

Filter permissions by property or value

	Principal ▲	Principal type ▼	Resource type ▼	Database ▼	Table ▼	Resource ▼
<input type="radio"/>	LF-GlueServiceRole	IAM role	Database	tpc	-	tpc
<input type="radio"/>	If-admin	IAM user	Database	tpc	-	tpc

# ワークフローを起動

- 画面左側よりBlueprintsを選択し、ワークフロー(tpc-workflow) が作成済であることを確認
- tpc-workflowを選択し、Actions->Startをクリック

**Workflows (0/1)**

Workflows are instances of ingestion blueprints in Lake Formation.

Find workflows

Actions

- Start
- Delete
- View graph

Use blueprint

< 1 > ⚙

Name	Created on	Last updated	Last run status
tpc-workflow	2021年10月18日(月) 4:45 UTC	2021年10月18日(月) 4:45 UTC	-

# tpc-workflowの実行について

- tpc-workflowの名前をクリックすると、詳細画面（下記）に遷移します
- Run IDをクリックするとGlueコンソールが開き、どのようなジョブが構成されたかを確認できます
- 今回の構成ではtpc-workflowの起動から実行完了までは25分程度の時間がかかります
- そのため今回は、事前にS3上に用意されているNYCTAXI表を手動で登録し、その表に対してLake Formationの権限設定を体験します
  - 他の方法としてGlueクローラーで登録する事も可能です

tpc-workflow

StartDeleteView graph

Workflow details

Name tpc-workflow	Last updated 2021年10月22日(金) 11:11 UTC
Last run status 🟢 COMPLETED	Created on 2021年10月22日(金) 11:11 UTC

Workflow runs (1)

Find Workflow runs

< 1 > ⚙

Name ▼	Started on ▲	Run ID ▼
tpc-workflow	2021年10月22日(金) 11:21 UTC	<a href="#">wr_a22115296f57bdc7113764acfd29b238ad512f6f917e586241eb2f6d994efd19</a>



# 補足：nyctaxiデータ (CSVファイル)

S3コンソールを開くと、データレイク用バケット (lf-data-lake-\*)以下に、**nyctaxiフォルダ**があり、その中にタクシーの走行に関するデータが記録された**CSVファイル**(tripdata-noheader.csv)が保存されている事が確認できます

※補足：lf-adminはs3:ListBucket等の権限を持たないため、画面の一部にエラーが出ますが今回の操作には問題ありません

The screenshot shows the Amazon S3 console interface for the bucket 'lf-data-lake-nyctaxi/'. The breadcrumb path is 'Amazon S3 > lf-data-lake-nyctaxi/'. The folder name 'nyctaxi/' is displayed at the top. A button 'S3 URI をコピー' is in the top right. Below the folder name, there are tabs for 'オブジェクト' (selected) and 'プロパティ'. The main content area is titled 'オブジェクト (1)' and contains a description of S3 objects. Below the description are buttons for 'アップロード', 'S3 URI をコピー', 'URL をコピー', 'ダウンロード', '開く', '削除', 'アクション', and 'フォルダの作成'. A search bar with the placeholder 'プレフィックスでオブジェクトを検索' and a 'バージョンの表示' toggle are also present. At the bottom, a table lists the objects:

<input type="checkbox"/>	名前	▲	タイプ ▼	最終更新日時 ▼	サイズ ▼	ストレージクラス ▼
<input type="checkbox"/>	tripdata-noheader.csv		csv	2021/10/22 06:06:11 PM JST	1.7 MB	スタンダード

# nyctaxi表を手動登録

- Lake Formationコンソールで画面左のTablesを選択し、“Create table”をクリック

Table details:

- Nameにnyctaxiと入力
- Databaseはtpcを選択  
(次ページに続く)

AWS Lake Formation > Tables > Create table

## Create table

**Table details**  
Create a table in the AWS Glue Data Catalog.

☒ **Table**  
Create a table in my account.

☐ **Resource link**  
Create a resource link to a shared table.

**Name**  
nyctaxi

Name may contain letters (A-Z), numbers (0-9), hyphens (-), or underscores (\_), and must be less than 256 characters long.

**Database**  
Table is contained within this database.

tpc

[Create database](#)

**Description - optional**  
Enter a description

Descriptions can be up to 2048 characters long.

# nyctaxi表を手動登録

## Data store:

- Specified path in my accountを選択
- Include pathで”Browse”を押してlf-data-lake...から始まるバケットの右側にある”>”をクリックし、nyctaxiパスを選択する

## Data format

- CSVを選択

(次ページに続く)

### Data store

Data is located in

- ☒ Specified path in my account  
☐ Specified path in another account

#### Include path

Path must be in the form s3://bucket/prefix. It must end with a slash (/) and not include any files.

s3://lf-data-lake- /nyctaxi/

Browse

### Data format

#### Classification

- ☐ Avro  
☒ CSV  
☐ JSON  
☐ XML  
☐ PARQUET  
☐ ORC

Choose the format of the data in your table.

#### Delimiter

Comma: ,

# nyctaxi表を手動登録

Schema:

- Upload Schemaボタンを押す
- ダイアログが開くので、配布ファイルのnyctaxi-schema.txtの内容をコピー＆ペーストし、Uploadボタンを押す（※Uploadボタンがグレースアウトして押せないように見える場合、一度Uploadボタンをクリックすると押せるようになります）
- 元の画面に戻るので、"Submit"を押すと、表が登録できます

The screenshot shows the 'Schema' management interface. On the left, there is a search bar labeled 'Find Columns' and a button labeled 'Upload Schema'. Below this, it says 'No available schema'. A blue arrow points from the 'Upload Schema' button to a modal dialog titled 'Upload schema'. The dialog has a close button (X) in the top right corner. Inside the dialog, there is a text input field with the placeholder text 'Paste a JSON array of column structures to create a schema. [Learn More](#)'. A blue arrow points to this input field, which contains a partial JSON array: 

```
[  
  {  
    "Name": "vendorid",  
    "Type": "bigint"  
  },  
  {
```

. At the bottom of the dialog, there are 'Cancel' and 'Upload' buttons. A blue arrow points to the 'Upload' button. Below the dialog, at the bottom of the page, there are 'Cancel' and 'Submit' buttons.

# 補足：If-adminへの 暗黙的な権限付与について

If-adminにはtpcデータベースへのCREATE TABLE権限をGrantしていないのに、表を手動で作成できました

これは、If-adminがtpcデータベースを作成した際に自動的にtpc内へのCREATE TABLE権限などが付与されたため

このようにLake Formationでは、一般的な利用概念に即して暗黙的な権限付与が実行されます

参考：

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

nyctaxi

Version 0 (Current version) ▾

Actions ▾

Compare versions

Drop

## Table details

Database tpc	Description -	Governance Disabled
Location s3://lf-data-lake- /nyctaxi/	Data format csv	Compaction Status -
Connection -	Last updated 2021年10月18日(月) 12:23 UTC	

► Advanced table properties

## Schema

Find Columns

#	Column Name ▾	Data type ▾	Partition key	Comment
1	vendorid	bigint	-	-
2	lpep_pickup_datetime	string	-	-
3	lpep_dropoff_datetime	string	-	-
4	store_and_fwd_flag	string	-	-
5	ratecodeid	bigint	-	-
6	pulocationid	bigint	-	-
7	dolocationid	bigint	-	-
8	passenger_count	bigint	-	-
9	trip_distance	double	-	-
10	fare_amount	double	-	-
11	extra	double	-	-



# 補足：nyctaxi表の定義

nyctaxiの定義は右図の通り

この内容は配布ファイルの  
nyctaxi-schema.txtでも確認  
可能（JSON形式で表定義の  
情報が格納されている）

#	Column Name	Data type
1	vendorid	bigint
2	lpep_pickup_datetime	string
3	lpep_dropoff_datetime	string
4	store_and_fwd_flag	string
5	ratecodeid	bigint
6	pulocationid	bigint
7	dolocationid	bigint
8	passenger_count	bigint
9	trip_distance	double
10	fare_amount	double
11	extra	double
12	mta_tax	double
13	tip_amount	double
14	tolls_amount	double
15	ehail_fee	string
16	improvement_surcharge	double
17	total_amount	double
18	payment_type	bigint
19	trip_type	bigint

# 表・列単位でアクセスを設定する

If-developerユーザに対し、表と列を名前で指定し、そのみ読み取り可能になるよう設定します

- 画面左からData lake permissionsを選択し、右側の”Grant”を押す
- Principals :
  - ”IAM users and roles”を選択
  - IAM UserのIf-developerを選択
- LF-Tags or catalog resources :
  - ”Named data catalog resources”を選択
  - Databaseにtpcを選択
  - Tablesにnyctaxiを選択(次ページに続く)

The screenshot shows the 'Grant' interface in the AWS IAM console. It is divided into two main sections: 'Principals' and 'LF-Tags or catalog resources'.

**Principals**

- Under 'IAM users and roles', the 'If-developer User' is selected in the dropdown menu.

**LF-Tags or catalog resources**

- Under 'Named data catalog resources', the 'tpc' database is selected in the dropdown menu.
- Under 'Tables - optional', the 'nyctaxi' table is selected in the dropdown menu.

Blue arrows point to the following elements:

- The 'IAM users and roles' radio button.
- The 'If-developer User' dropdown selection.
- The 'Named data catalog resources' radio button.
- The 'tpc' database dropdown selection.
- The 'nyctaxi' table dropdown selection.

# 表・列単位でアクセスを設定する

Table permissionsの中  
のTable permissions  
で”Select”のみにチェッ  
クを入れる

(次ページに続く)

**Table permissions**

Table permissions  
Choose specific access permissions to grant.

<input checked="" type="checkbox"/> Select	<input type="checkbox"/> Insert	<input type="checkbox"/> Delete
<input type="checkbox"/> Describe	<input type="checkbox"/> Alter	<input type="checkbox"/> Drop

**Grantable permissions**  
Choose the permission that may be granted to others.

<input type="checkbox"/> Select	<input type="checkbox"/> Insert	<input type="checkbox"/> Delete
<input type="checkbox"/> Describe	<input type="checkbox"/> Alter	<input type="checkbox"/> Drop

☐ Super  
This permis  
the left, an

☐ Super  
This permis  
permissions  
permissions

# 表・列単位でアクセスを設定する

列を限定して閲覧を許可します

Data permissions:

- Column-based accessを選択
- Include columnsを選択
- Select columnsで、vendorid、lpep\_pickup\_datetime、lpep\_dropoff\_datetime、passenger\_count、trip\_distanceを選択

“Grant”を押す

If-developerへのアクセス設定が完了しました

The screenshot shows the 'Data permissions' configuration window. It has two main sections: 'Data permissions' and 'Choose permission filter'. In the 'Data permissions' section, the 'Column-based access' radio button is selected. In the 'Choose permission filter' section, the 'Include columns' radio button is selected. Below this, the 'Select columns' dropdown menu is open, showing a list of columns: vendorid (bigint), lpep\_pickup\_datetime (string), lpep\_dropoff\_datetime (string), passenger\_count (bigint), and trip\_distance (double). At the bottom, there is a 'Grantable permissions' section with a 'Select' checkbox. Blue arrows point to the 'Column-based access' radio button, the 'Include columns' radio button, the 'Select columns' dropdown menu, and the 'Grant' button at the bottom right.

**Data permissions**

☐ All data access  
Grant access to all data without any restrictions.

☒ Column-based access  
Grant data access to specific columns only.

**Choose permission filter**  
Choose whether to include or exclude columns.

☒ Include columns  
Grant permissions to access specific columns.

☐ Exclude columns  
Grant permissions to access all but specific columns.

**Select columns**

Choose one or more columns

vendorid ×  
bigint

lpep\_pickup\_datetime ×  
string

lpep\_dropoff\_datetime ×  
string

passenger\_count ×  
bigint

trip\_distance ×  
double

**Grantable permissions**  
Choose the permission that may be granted to others.

☐ Select

Cancel Grant

# LF-Tagベースで権限を設定する


lf-campaign\_managerユーザーに対しては、LF-Tagベースで権限を設定します

まずLF-Tagを定義し、それを対象リソースに付与、その後ユーザーにタグを付与するという手順で設定します


- 画面左にあるLF-Tagsをクリックし、ADD LF-tagを押す
- Keyにはgroupと入力
- Valuesにはdeveloper, campaign, analystをそれぞれAddする
- Add LF-tagを押す

**Add LF-Tag** [Learn More](#) ×

LF-Tags have a key and one or more values that can be associated with data catalog resources. Tables automatically inherit from database LF-tags, and columns inherit from table LF-tags.  
Example: Key = Confidentiality | Values = private, sensitive, public

**Key** 

Key string must be less than 128 characters long, and cannot be changed once LF-tag is created.


**Values**  
Type a single value and select [Enter] or specify multiple values separated by commas. 

developer ×

campaign ×

analyst ×

Enter up to 15 values; each value must be less than 256 characters long.

Cancel Add LF-tag 

# LF-Tagを列に付与する

作成したタグをNYCTAXI表の中のいくつかの列に付与します

- 左側Tablesをクリックし、nyctaxi表をクリック（名前の部分をクリック）
- 画面下部右側のEdit schemaをクリック

**Tables (25)**

Find table by properties

	Name	Database	Owner account ID	Shared resou...	Shared resource ow...	Location
<input type="radio"/>	nyctaxi	tpc	-	-	-	s3://lf-da...

**Schema**

Find Columns

#	Column Name	Data type	Partition key	Comment	LF-Tags
1	vendorid	bigint	-	-	1

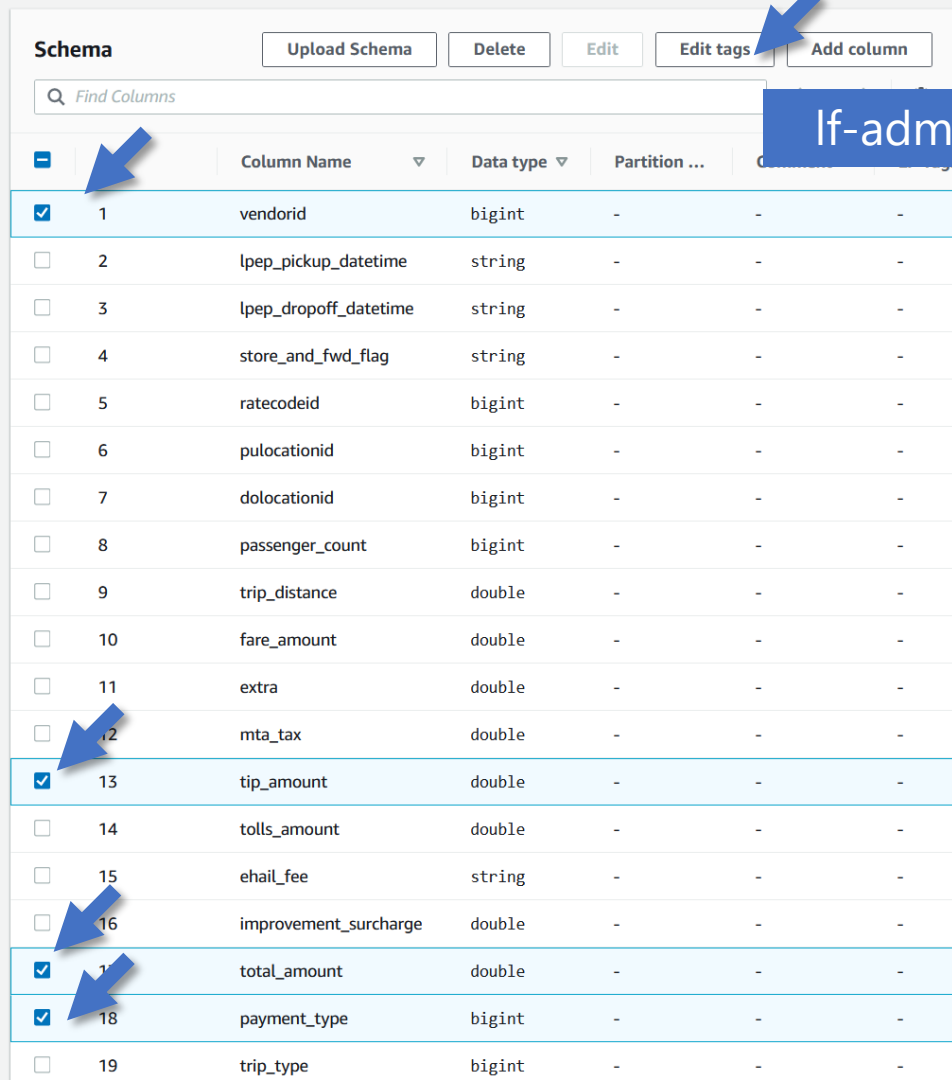
# LF-Tagを列に付与する

右図のように、

- vendorid
- tip\_amount
- total\_amount
- payment\_type

にチェックをいれ、  
上部の“Edit tags”を押す

※AWSが提供する環境の場合、ここで“Unknown error”が表示される事がありますが、そのまま進めても問題ありません



	Column Name	Data type	Partition ...	
<input checked="" type="checkbox"/>	1 vendorid	bigint	-	-
<input type="checkbox"/>	2 lpep_pickup_datetime	string	-	-
<input type="checkbox"/>	3 lpep_dropoff_datetime	string	-	-
<input type="checkbox"/>	4 store_and_fwd_flag	string	-	-
<input type="checkbox"/>	5 ratecodeid	bigint	-	-
<input type="checkbox"/>	6 pulocationid	bigint	-	-
<input type="checkbox"/>	7 dolocationid	bigint	-	-
<input type="checkbox"/>	8 passenger_count	bigint	-	-
<input type="checkbox"/>	9 trip_distance	double	-	-
<input type="checkbox"/>	10 fare_amount	double	-	-
<input type="checkbox"/>	11 extra	double	-	-
<input type="checkbox"/>	12 mta_tax	double	-	-
<input checked="" type="checkbox"/>	13 tip_amount	double	-	-
<input type="checkbox"/>	14 tolls_amount	double	-	-
<input type="checkbox"/>	15 ehail_fee	string	-	-
<input type="checkbox"/>	16 improvement_surcharge	double	-	-
<input checked="" type="checkbox"/>	17 total_amount	double	-	-
<input checked="" type="checkbox"/>	18 payment_type	bigint	-	-
<input type="checkbox"/>	19 trip_type	bigint	-	-

# LF-Tagを列に付与する

- “Assign new LF-Tag”を押し、keyはgroup、Valuesにはcampaignを選択して”Save”を押す
- 元の画面に戻るの  
で、”Save as new  
version”を押す

**Edit LF-Tags: 4 columns** [Learn More](#) ×

i Showing only LF-tags that are shared by all selected columns.

**LF-Tags**

After they are associated with catalog resources, LF-Tags allow you to create scalable permissions.

There are no inherited LF-Tags associated with the resource.

Assigned keys	Values	
<input type="text" value="group"/> <span>×</span>	<input type="text" value="campaign"/> ▼	<input type="button" value="Remove"/>
<input type="button" value="Assign new LF-Tag"/>		
<p>You can add 49 more LF-tags.</p>		
		<input type="button" value="Cancel"/> <input type="button" value="Save"/>



# LF-Tagをユーザに付与する

LF-Tagのリソースへの付与が完了したので、ユーザにLF-Tagへの権限を付与します

画面左のData lake permissionsをクリックし、“Grant”を押す

Principals:

- IAM users and rolesを選択
- lf-campaign-managerを選択

LF-Tags or catalog resources

- Resources mached by LF-Tagsを選択
- Add LF-Tagを押す
- Keyにgroupを選択
- Valuesにcampaignのみを選択

(次ページに続く)

## Grant data permissions

### Principals



#### IAM users and roles

Users or roles from this AWS account.



#### SAML users and groups

SAML users and group or QuickSight ARNs.



#### External accounts

AWS accounts or AWS organizations outside of this account.

### IAM users and roles

Add one or more IAM users or roles.

Choose IAM principals to add

lf-campaign-manager X  
User

### LF-Tags or catalog resources



#### 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

Manager permissions for specific databases or tables, in addition to fine-grained data access.

### Key

Q group X

Add LF-Tag

### Values

Choose LF-tag values

☐ analyst

☒ campaign

☐ developer

Remove

# LF-Tagをユーザに付与する

Database permissions: は変更なし

Table permissions:

- Table permissionでSelectを選択
- “Grant”を押す

**Table permissions**

Table permissions  
Choose specific permissions to grant.

☒ Select ☐ Insert ☐ Delete  
☐ Describe ☐ Alter ☐ Drop

**Grantable permissions**  
Choose the permission that may be granted to others.

☐ Select ☐ Insert ☐ Delete  
☐ Describe ☐ Alter ☐ Drop

☐ Super  
This permission is the union of all the individual permissions to the left, and supersedes them.

☐ Super  
This permission allows the principal to grant any of the permissions to the left, and supersedes those grantable permissions.

Cancel Grant

# tpcデータベースへのdescribe権限を付与する

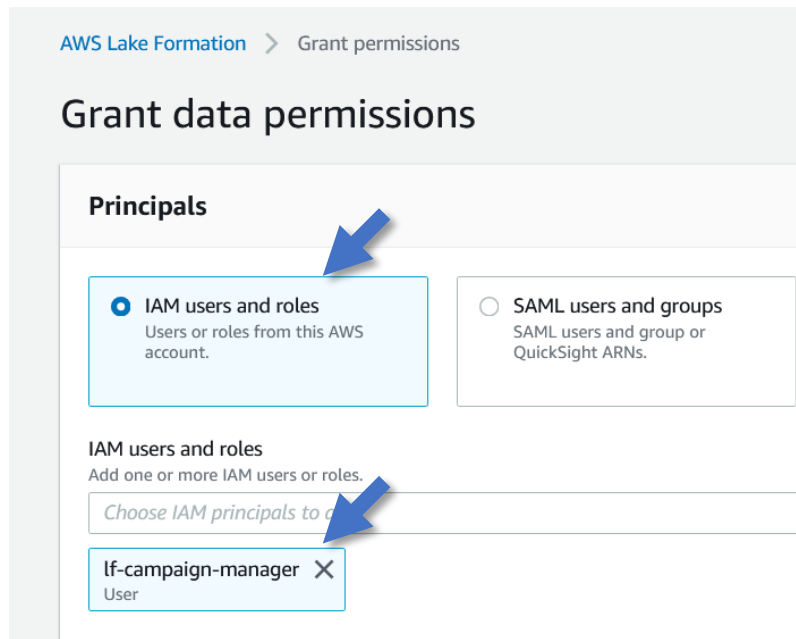
LF-Tagで列に対してアクセスができるよう設定しましたが、後述のAthenaを利用する際にDBから表一覧を得る権限(describe)も必要になるため、追加で権限を設定します

画面左のData lake permissionsをクリックし、“Grant”を押す

Principals:

IAM users and rolesを選択  
lf-campaign-managerを選択

(次ページに続く)



# tpcデータベースへのdescribe権限を付与する

lf-admin

LF-Tags or catalog resources:

- Named data catalog resourcesを選択
- Databasesでtpcを選択

Database permissions

- Describeを選択

“Grant”を押す

これでlf-developerと、lf-campaign-managerへの権限設定は完了です

The screenshot shows the 'lf-admin' interface for granting permissions. It is divided into two main sections: 'LF-Tags or catalog resources' and 'Database permissions'.

**LF-Tags or catalog resources**

- Under 'Resources matched by LF-Tags (recommended)', there is a description: 'Manage permissions indirectly for resources or data matched by a specific set of LF-Tags.'
- Under 'Named data catalog resources', there is a description: 'Manage permissions for specific databases or tables, in addition to fine-grained data access.' A blue arrow points to this section.
- Under 'Databases', there is a dropdown menu labeled 'Choose databases'. The 'tpc' database is selected and highlighted with a blue box. A blue arrow points to this dropdown.
- There is a 'Load more' button next to the database dropdown.
- Under 'Tables - optional', there is a dropdown menu labeled 'Choose tables' and a 'Load more' button.

**Database permissions**

- Under 'Database permissions', there is a description: 'Choose specific access permissions to grant.'
- There are three checkboxes: 'Create table', 'Alter', and 'Drop'. The 'Describe' checkbox is checked with a blue checkmark. A blue arrow points to this checkbox.
- Under 'Grantable permissions', there is a description: 'Choose the permission that may be granted to others.'
- There are three checkboxes: 'Create table', 'Alter', and 'Drop'. The 'Describe' checkbox is checked.
- There are two 'Super' permission options, each with a description: 'This permission is the union of all the individual permissions to the left, and supersedes them.'
- At the bottom right, there are two buttons: 'Cancel' and 'Grant'. A blue arrow points to the 'Grant' button.

# Lake Formation利用ユーザで 動作確認

(If-developer, If-campaign-manager)

# If-developerユーザでのログイン

リージョンに注意！



If-adminからサインアウトし、  
If-developerでログインしなおします（ログインのURLなどはIf-admin  
ログイン時のページを参照。パスワードはIf-adminと同じ）

画面右上のユーザ名がIf-developer @ ...になっていることを確認

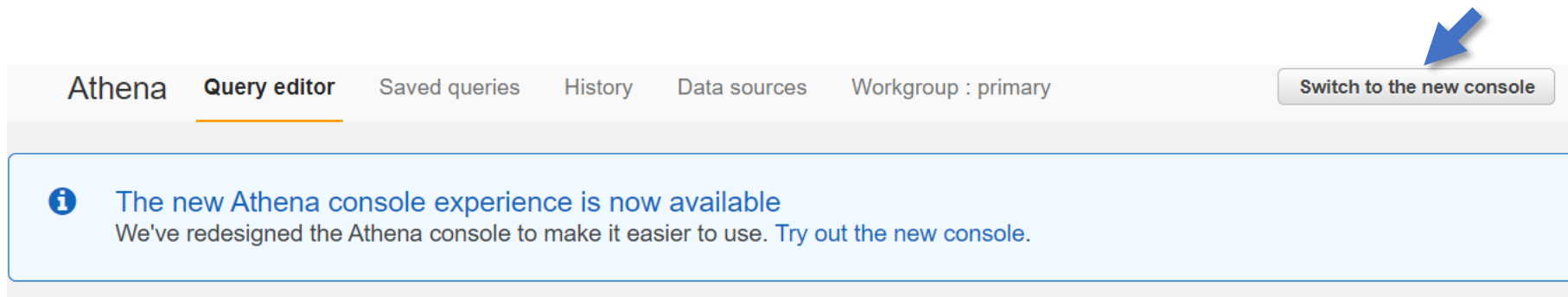
リージョンを必ず確認し、「バージニア北部」を選択してください

確認できたらAthenaコンソールを開きます

# 補足：Athenaの旧コンソールについて

Athenaのコンソールは現在、新しいバージョンに切り替わっている最中であり、ユーザ切り替えのタイミングで旧Athenaコンソールが表示される場合があります

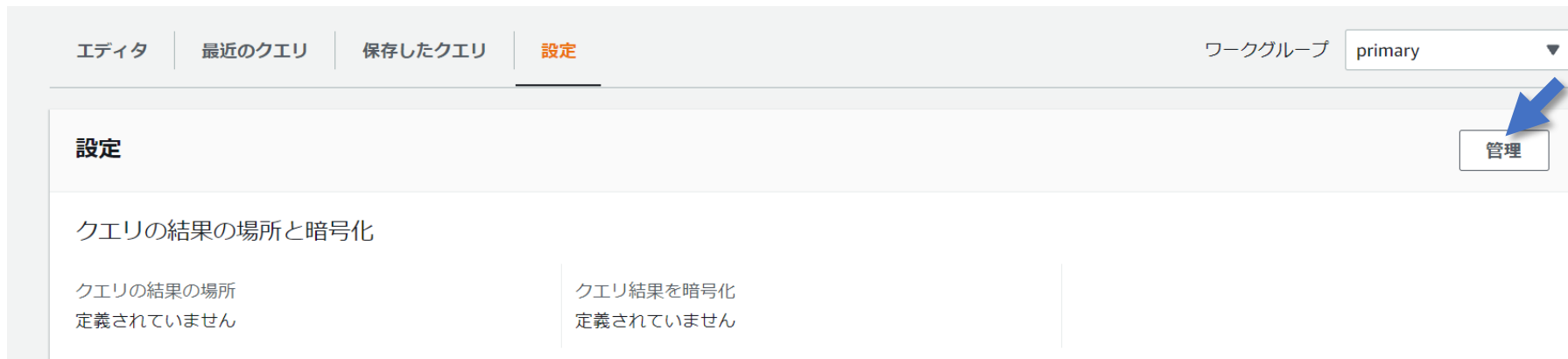
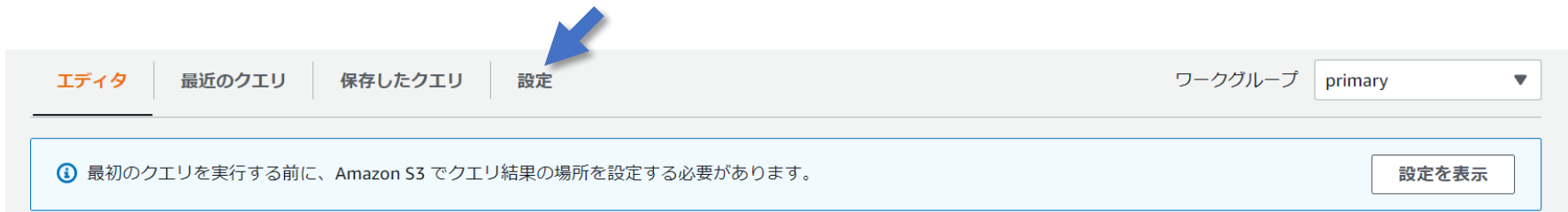
この場合、以下の画面が表示されるので、“Switch to the new console”を選んで新コンソールに切り替えてください



# Athenaのクエリ結果保存バケットをセット

Athenaを利用開始時に、クエリ結果を保存するS3 URLを指定する必要があります

- 画面左サイドのメニューより”クエリエディタ”をクリック
- “設定”を押す => 設定タブ内の“管理”を押す





# Athenaのクエリ結果保存バケットをセット

- CloudFormationの”出力”にAthenaQueryResultLocationで記載されていたURL(メモしたもの) をクエリ結果の場所にペーストして、”保存”
- エディタを押して、エディタ画面に戻る

Amazon Athena > クエリエディタ > 設定を管理

### 設定を管理

#### クエリの結果の場所と暗号化

クエリ結果の場所

☐ クエリ結果を暗号化

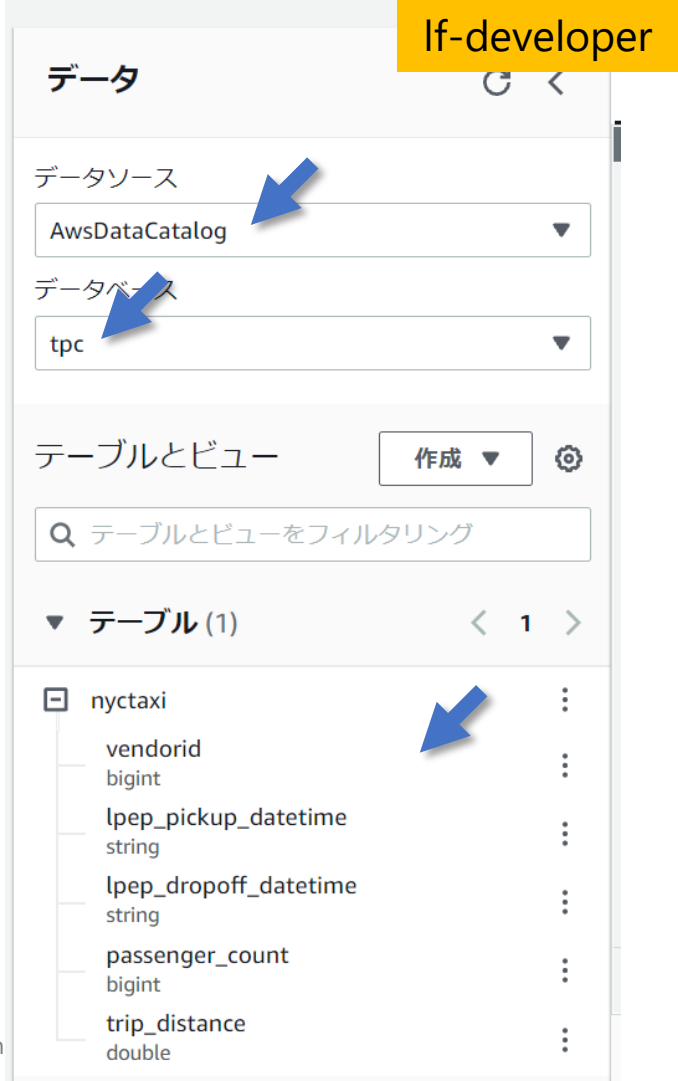
[表示](#) [S3を参照](#)

[キャンセル](#) [保存](#)

# If-developerの権限を確認

クエリエディタ左側のペインで  
データソース：AwsDataCatalog  
データベース：tpc  
になっている事を確認

テーブルにはnyctaxiがあるので、  
展開すると右図のようにnyctaxi表  
のうち許可した5列のみ見えている  
事が確認できる



# If-developerの権限を確認

nyctaxiのメニューから”テーブルをプレビュー”を押すか、  
クエリエディタで `SELECT * FROM nyctaxi LIMIT 10;`と書いて”実行”をクリックします  
`SELECT *`で前列を指定しても、許可されていない列は現れないことが分かります

The screenshot shows the Amazon Redshift console interface. On the left, the 'nyctaxi' table is selected under the 'tpc' database. The main area displays the table's schema and a preview of its data. A context menu is open over the table, with a blue arrow pointing to the 'テーブルをプレビュー' (Preview Table) option.

vendorid	lpep_pickup_datetime	lpep_dropoff_datetime	passenger_count	trip_distance
1	2017-01-01 00:01:15	2017-01-01 00:11:05	1	1.71
-01-01	00:03:34	2017-01-01 00:09:00	1	1.44
-01-01	00:04:02	2017-01-01 00:12:55	5	3.45
-01-01	00:01:40	2017-01-01 00:14:23	1	2.11
-01-01	00:00:51	2017-01-01 00:18:55	1	2.76
-01-01	00:00:28	2017-01-01 00:13:31	1	4.14

# If-developerの権限を確認

SELECT extra FROM  
nyctaxi;

等、許可されていない列  
をクエリしてみるとエ  
ラーになることが確認で  
きます

(列が存在しないような  
動作になっている)

The screenshot displays the Amazon Redshift console interface. On the left, the 'データ' (Data) section shows the 'nyctaxi' table selected from the 'tpc' database. The table schema lists columns: vendorid (bigint), lpep\_pickup\_datetime (timestamp), lpep\_dropoff\_datetime (timestamp), passenger\_count (bigint), and trip\_distance (double). The main area shows a query editor with the text 'SELECT extra FROM nyctaxi LIMIT 10;'. Below the query, a red error message is displayed: 'SYNTAX\_ERROR: line 1:8: Column 'extra' cannot be resolved'. The message explains that the column 'extra' is not found in the 'tpc' database and provides a link to the Redshift documentation for more information.

データ

データソース  
AwsDataCatalog

データベース  
tpc

テーブルとビュー 作成 ▼ ⚙️

Q テーブルとビューをフィルタリング

▼ テーブル (1) < 1 >

- nyctaxi
  - vendorid (bigint)
  - lpep\_pickup\_datetime (timestamp)
  - lpep\_dropoff\_datetime (timestamp)
  - passenger\_count (bigint)
  - trip\_distance (double)

クエリ 1 ✖️ | クエリ 2 ✔️

1 SELECT extra FROM nyctaxi LIMIT 10;

行 1、列 36

Run again キャンセル 名前を付けて保存 クリア

✖️ 失敗 キュー内の時間: 0.195 秒

✖️ SYNTAX\_ERROR: line 1:8: Column 'extra' cannot be resolved  
このクエリは、クエリで修飾されていない限り、「tpc」データベースに存在しない列 'extra' を参照しています。  
[フォーラム](#) に投稿するか、クエリ ID: 44160513-e422-4840-863d-c5c...  
にお問い合わせください。

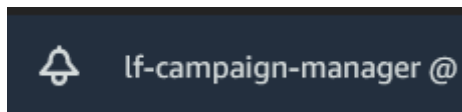
# If-developerの権限を確認

DROP TABLE nyctaxi; を実行します

DROP権限が無い（SELECTしか許可されていないため）  
削除できないことが分かります（AccessDeniedException）

The screenshot shows the AWS Glue console interface. On the left, the 'データ' (Data) sidebar is visible, showing the 'データソース' (Data Source) as 'AwsDataCatalog' and the 'データベース' (Database) as 'tpc'. Under 'テーブルとビュー' (Tables and Views), the table 'nyctaxi' is listed. The main area shows a SQL query editor with the statement 'DROP TABLE nyctaxi;'. Below the editor, the 'Run again' button is highlighted. The execution status is '失敗' (Failed). The error message is displayed in a red box: 'FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.qlexec.DDLTask. MetaException(message:Insufficient Lake Formation permission(s): Required Drop on nyctaxi (Service: AmazonDataCatalog; Status Code: 400; Error Code: AccessDeniedException; Request ID: 8bd831bf-7e97-4c5e-9419-e786aeb07f44; Proxy: null))'. The error message is followed by a link to the forum and a link to customer support.

# If-campaign-managerに切り替える リージョンに注意！



If-developerからサインアウトし、  
If-campaign-managerでログインしなおします

画面右上のユーザ名がIf-campaign-manager @ ...になっていることを確認

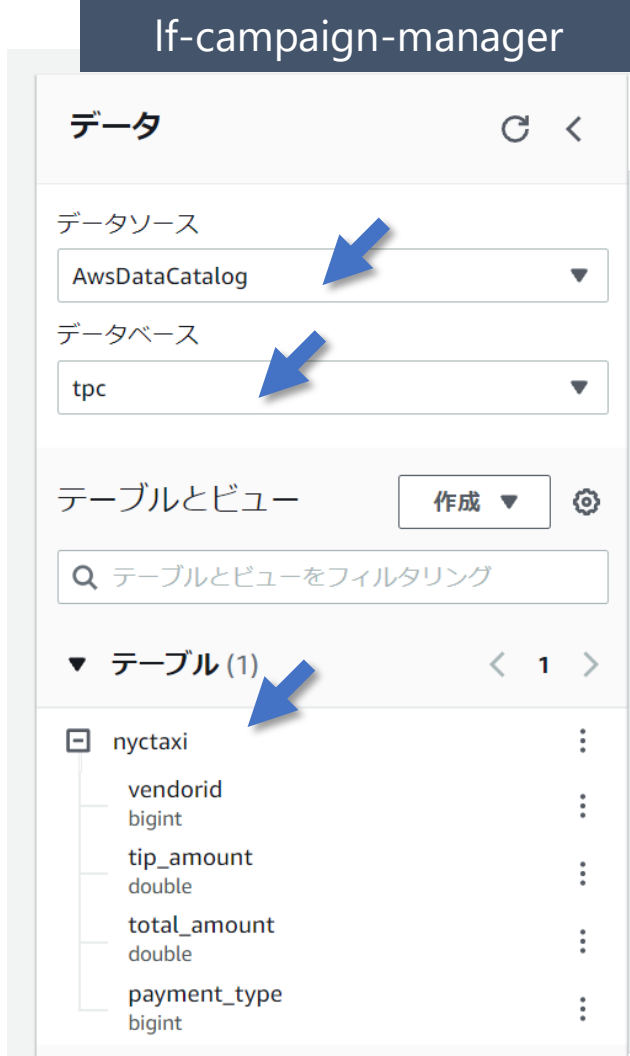
リージョンを必ず確認し、「バージニア北部」を選択してください

Athenaコンソールを開き、クエリ結果を保存するS3 URLを設定します  
(If-developerの時と同じ方法ですので、そちらを参照してください)

# If-campaign-managerの権限を確認

Athenaクエリエディタ左側のペインで  
データソース：AwsDataCatalog  
データベース：tpc  
になっている事を確認

テーブルにはnyctaxiがあるので、展開すると右図のようにnyctaxi表のうち**LF-TAG**を付けた4列のみ見えている事が確認できる



# lf-campaign-managerの権限を確認

テーブルのプレビュー、  
もしくはSELECT \*  
FROM nyctaxi LIMIT  
10;

を実行すると、4列の  
み読み取り可能である  
ことが確認できます

The screenshot displays the Amazon Redshift console interface. On the left sidebar, the 'データ' (Data) section is active, showing the 'nyctaxi' table under the 'tpc' database. The main panel shows a 'New query 1' execution. The query 'SELECT \* FROM nyctaxi LIMIT 10;' has been executed successfully, as indicated by the green '完了' (Completed) status bar. The execution details show a query time of 0.171 seconds and a scan time of 1.044 seconds. The results are displayed in a table with 4 columns: 'vendorid', 'tip\_amount', 'total\_amount', and 'payment\_type'. The table contains 10 rows of data.

vendorid	tip_amount	total_amount	payment_type
2	0.0	9.8	2
2	0.0	7.8	2
2	2.66	15.96	1
2	0.0	11.8	2
2	0.0	12.8	2



# lf-campaign-managerの権限を確認

SELECT extra  
FROM nyctaxi; など、許可されていない列をクエリしてみるとエラーが変えることが確認できます

同様にDROP  
TABLE nyctaxi;も  
実行できないことが確認できます

The screenshot displays the Amazon Redshift console interface. On the left, the 'データ' (Data) section shows the 'データソース' (Data Source) as 'AwsDataCatalog' and the 'データベース' (Database) as 'tpc'. Below this, the 'テーブルとビュー' (Tables and Views) section shows a search bar and a list of tables, including 'nyctaxi' with its columns: 'vendorid', 'bigint', 'tip\_amount', 'double', 'total\_amount', 'double', and 'payment type'. On the right, the 'New query 1' section shows the SQL query: 'SELECT extra FROM nyctaxi LIMIT 10;'. Below the query, it indicates '行 1、列 13' (Line 1, Column 13). The 'Run again' button is highlighted in orange. A red error message box at the bottom right states: 'SYNTAX\_ERROR: line 1:8: Column 'extra' cannot be resolved. このクエリは、クエリで修飾されていない限り、「tpc フォーラム」に投稿するか、クエリ ID: ddb0dc6b-74... にお問い合わせください。' (This query, unless qualified in the query, should be posted to 'tpc Forum' or contact support with query ID: ddb0dc6b-74...).

## 補足：Blueprint ワークフローの結果を確認する

時間に余裕がある場合はlf-adminのAWSマネジメントコンソールに戻り、Blueprintの結果を確認してください

- Lake FormationコンソールのBlueprintを選択しtpc-workflowがCOMPLETEDになっている事を確認
- Lake FormationコンソールのTablesをクリックすると多くの表が登録されている事が確認できます
- Table一覧から任意の名前をクリックし、LocationやSchemaを確認してください  
※アンダースコア(\_)から始まる名前の表は、Blueprintが管理に使用するためのものですので、アンダースコアが付いていない表を選択してください
- S3コンソールを開き、lf-data-lake...バケット以下を確認
  - 指定したPrefix (dl)が名前の先頭についたフォルダが多数作成されていることが確認できます
  - どれか一つをクリックしてその下のディレクトリ構造やParquetファイルが存在する事を確認してください

# この後の進め方と クリーンアップ

# この後の進め方について

ここで今回のワークショップは終了です

ただし、この環境を使って <https://lakeformation.workshop.aws> の続きの内容を実施することが可能です（ただし一部修正した部分があるため、挙動や実行にかかる時間が異なる部分があるかもしれません）

続きを実施する場合は、費用に注意してください。EC2とRDSは必要になるまでサスペンドしておくで費用を抑える事が可能です

ワークショップを終える方はリソースのクリーンアップをしてください。  
クリーンアップはAWS Adminユーザで実施します  
(次ページより)





# クリーンアップ


- **AWS Admin**のブラウザに戻ってください
- S3コンソールでlf-data-lake-\*バケットを選択し、「空にする」を押して、バケット空にします
- 同様にlf-workshop-\*から始まるバケットを空にします

※ 2つのS3バケットを空にしておかないと、CloudFormationスタックの削除時にエラーになります



- CloudFormationコンソールで、最初の実行したスタックを削除します
- スタックの削除には約6分間かかりますが、削除完了を待たずにBlueprint削除等、次ページからのステップに進んでいただいても問題ありません



**バケット (2) 情報**   ARN をコピー  空にする  削除  バケットを作成

バケットは S3 に保存されたデータのためのコンテナです。 [詳細](#) 

🔍 バケットを名前を検索

名前 ▲	AWS リージョン ▼	アクセス ▼	作成日 ▼
 <a href="#">lf-data-lake-</a>	米国東部 (バージニア北部) us-east-1	<a href="#">オブジェクトは公開することができます</a>	2021/10/18 11:35:33 AM JST
 <a href="#">lf-workshop-</a>	米国東部 (バージニア北部) us-east-1	<a href="#">オブジェクトは公開することができます</a>	2021/10/18 11:35:33 AM JST

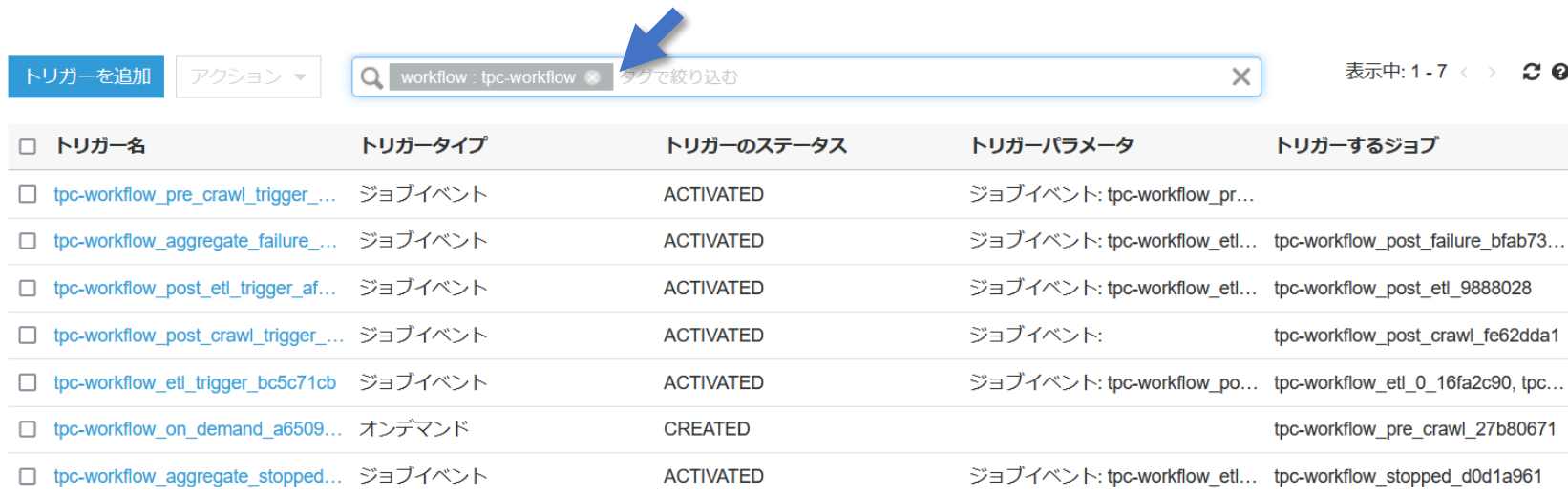
# クリーンアップ

CloudFormationスタックの削除によって...

- S3バケット、IAM User/Role、EC2、RDS、VPC関連のリソースは削除されています
- Lake Formationでの定義とBlueprint利用時に作られたGlueリソースやEC2 Key Pairは、CloudFormationで作成していないため、削除されません  
=>これをこの後のステップで手動削除します

# Lake FormationのBlueprintを削除

- Lake Formationコンソールに移動しBlueprints-> tpc-workflowを選択してDeleteします（これによりGlue側のtpc-workflowも消えます）
- Glueコンソールを開いて画面左からトリガーをクリックし、検索窓にworkflow:tpc-workflowを入れてフィルタをかけると、消すべき対象のトリガーのみに絞れるので、トリガーを選択し、アクションから削除を選択します
- ジョブ、クローラーについても同様に削除します（※クローラーは1つのみです。ジョブは複数ジョブを選択してGUIで消す方法はないため、1つずつ削除する必要があります）



トリガーを追加    アクション ▼    workflow : tpc-workflow    タグで絞り込む    表示中: 1 - 7 < > ↺ ⓘ

トリガー名	トリガータイプ	トリガーのステータス	トリガーパラメータ	トリガーするジョブ
<input type="checkbox"/> tpc-workflow_pre_crawl_trigger_...	ジョブイベント	ACTIVATED	ジョブイベント: tpc-workflow_pr...	
<input type="checkbox"/> tpc-workflow_aggregate_failure_...	ジョブイベント	ACTIVATED	ジョブイベント: tpc-workflow_etl...	tpc-workflow_post_failure_bfab73...
<input type="checkbox"/> tpc-workflow_post_etl_trigger_af...	ジョブイベント	ACTIVATED	ジョブイベント: tpc-workflow_etl...	tpc-workflow_post_etl_9888028
<input type="checkbox"/> tpc-workflow_post_crawl_trigger_...	ジョブイベント	ACTIVATED	ジョブイベント:	tpc-workflow_post_crawl_fe62dda1
<input type="checkbox"/> tpc-workflow_etl_trigger_bc5c71cb	ジョブイベント	ACTIVATED	ジョブイベント: tpc-workflow_po...	tpc-workflow_etl_0_16fa2c90, tpc...
<input type="checkbox"/> tpc-workflow_on_demand_a6509...	オンデマンド	CREATED		tpc-workflow_pre_crawl_27b80671
<input type="checkbox"/> tpc-workflow_aggregate_stopped...	ジョブイベント	ACTIVATED	ジョブイベント: tpc-workflow_etl...	tpc-workflow_stopped_d0d1a961

# Lake FormationのDBと表を削除する

Lake Formationコンソールに戻ります  
削除の操作をするために、今ログインしているAWS AdminにTPCデータベースへの権限を与えます

- Data lake permissionsをクリックし、右上の”Grant”をクリック
- IAM users and roles : ログインしているIAMユーザを選択 (AWSの用意したAWS環境で実施している場合は TeamRoleロールを選択)
- Named data catalog resourcesを選択し、databaseにtpcを選択
- Database permissions: 右図のように全部チェックして”Grant”を押す

## Database permissions

### Database permissions

Choose specific access permissions to grant.

- ☒ Create table    ☒ Alter    ☒ Drop  
☒ Describe

### Grantable permissions

Choose the permission that may be granted to others.

- ☐ Create table    ☐ Alter    ☐ Drop  
☐ Describe



# Lake FormationのDBと表を削除する

Lake Formationコンソールで以下のように削除する

- "Databases"から、データベース"tpc"を削除(テーブルの定義も合わせて削除される)
- "Data lake locations"から、S3パスが"s3://lf-data-lake\*"をRemoveする (※すでに存在しないリソースがあるというエラーが出る場合がありますが、Remove可能です)
- LF-TagsでgroupタグをDelete
- Administrative roles and tasksで"Choose administrator"を選択し、lf-adminを外してSave
  - 不要ならログインしている自分自身(AWS Admin)も外す

# EC2 Key PairとCloudWatch Logsの削除

- CloudWatchの管理画面で、画面左の「ログ」から「ロググループ」を選択し、/aws-glue/から始まるロググループを選択して削除します
  - ただし、すでに同リージョンでGlueを利用中の場合は、間違って既存のログを消さないようにご注意ください。ログは5GBまで無料枠があります
- 今回のワークショップのためにKey Pair作成した場合は、EC2の管理画面で作成したKey Pairを削除します
  - Key Pairは費用が発生しませんので、消さずに維持していただいても問題ありません

以上でクリーンアップは終了です。

おつかれさまでした！

# 参考資料

- AWS Lake Formation ドキュメントと関連動画  
以下にまとまっていますので、まずこちらを確認ください
  - <https://aws.amazon.com/jp/lake-formation/resources/>
- AWS Lake Formation Workshop (英語)
  - <https://lakeformation.workshop.aws/>

# 内容についての注意点

- 本資料では2021年11月19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。
- 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.