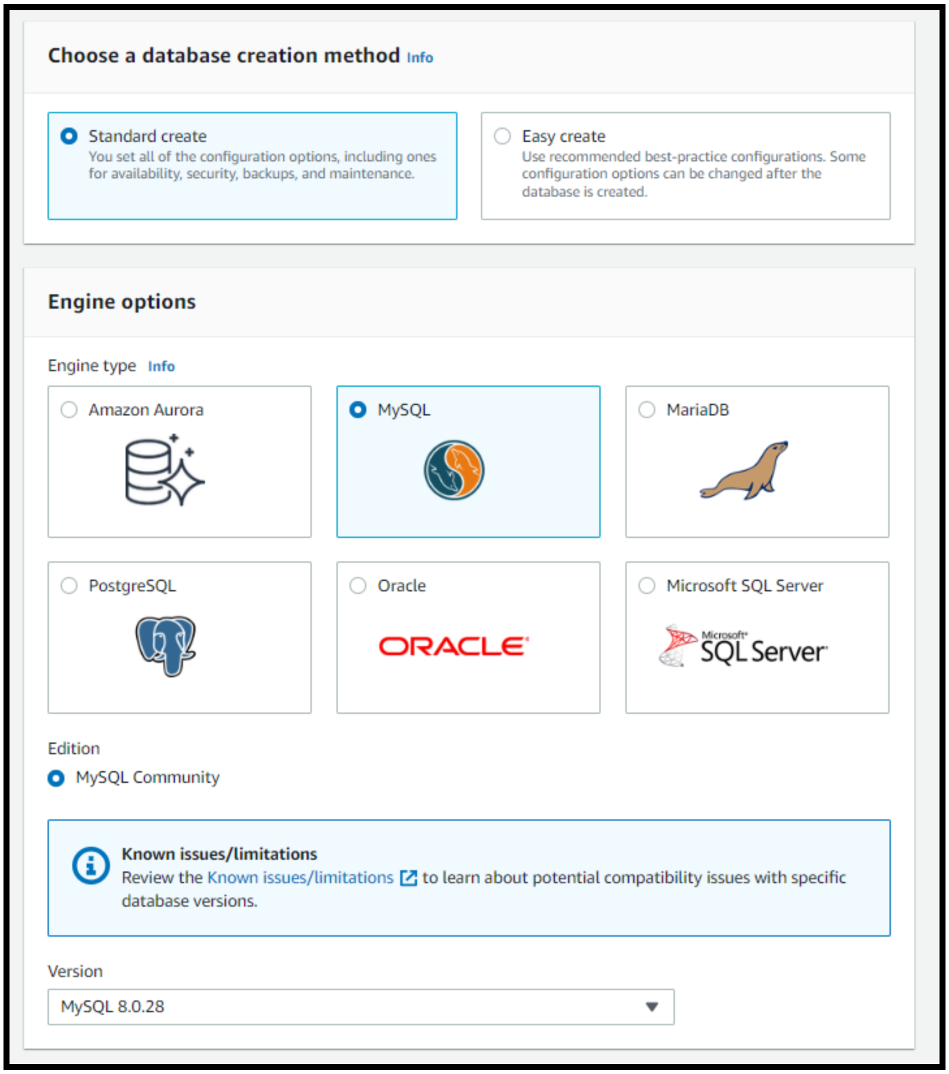
Data Lake Proof of Concept – Detailed Guide

By Carlos Cruz Mejia

Step by Step

1. Data Sources – MySQL

First of all, we need to set up our data source. We will proceed with a MySQL database with a standard configuration for this example. Make sure to select the free tier to keep costs as low as possible.



Choose a database creation method Info

● Standard create
You set all of the configuration options, including ones for availability, security, backups, and maintenance.

○ Easy create
Use recommended best-practice configurations. Some configuration options can be changed after the database is created.

Engine options

Engine type Info

○ Amazon Aurora

● MySQL

○ MariaDB

○ PostgreSQL

○ Oracle

○ Microsoft SQL Server

Edition

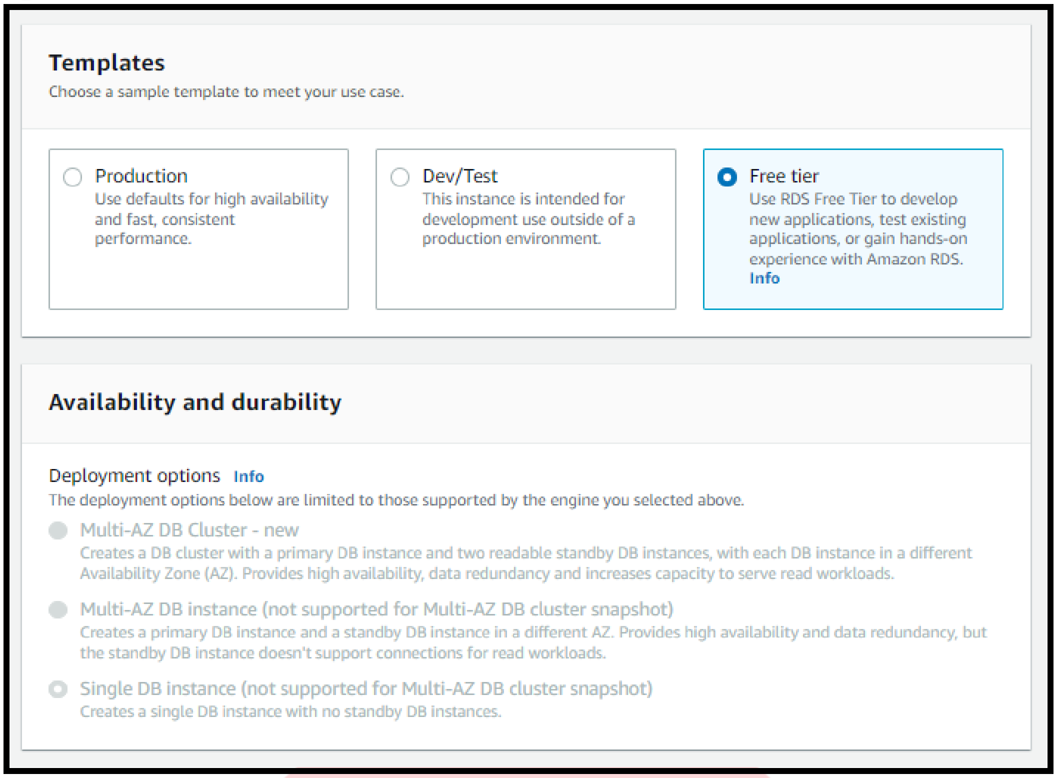
● MySQL Community

Known issues/limitations
Review the Known issues/limitations to learn about potential compatibility issues with specific database versions.

Version

MySQL 8.0.28

Then we must select the free tier template.



Templates

Choose a sample template to meet your use case.

○ Production
Use defaults for high availability and fast, consistent performance.

○ Dev/Test
This instance is intended for development use outside of a production environment.

● Free tier
Use RDS Free Tier to develop new applications, test existing applications, or gain hands-on experience with Amazon RDS.
Info

Availability and durability

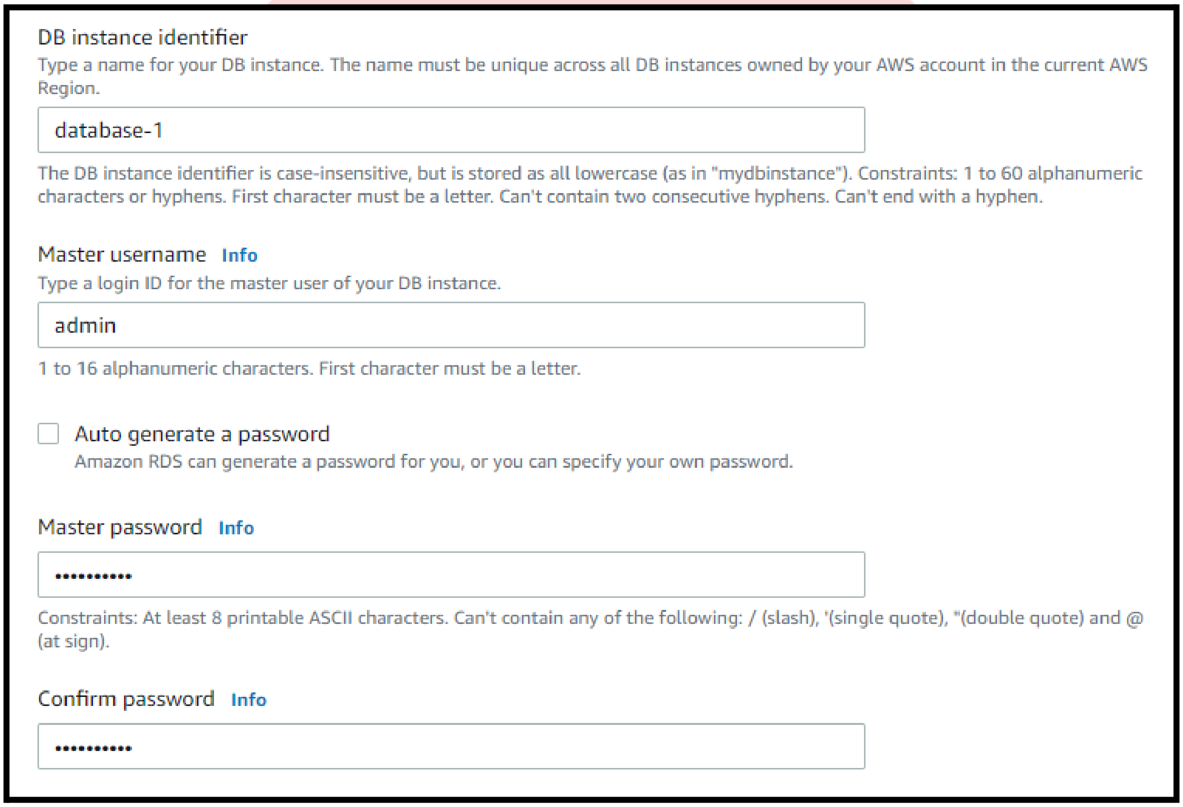
Deployment options Info
The deployment options below are limited to those supported by the engine you selected above.

○ Multi-AZ DB Cluster - new
Creates a DB cluster with a primary DB instance and two readable standby DB instances, with each DB instance in a different Availability Zone (AZ). Provides high availability, data redundancy and increases capacity to serve read workloads.

○ Multi-AZ DB instance (not supported for Multi-AZ DB cluster snapshot)
Creates a primary DB instance and a standby DB instance in a different AZ. Provides high availability and data redundancy, but the standby DB instance doesn't support connections for read workloads.

● Single DB instance (not supported for Multi-AZ DB cluster snapshot)
Creates a single DB instance with no standby DB instances.

We keep all default configurations in the instance identifier and set our master password.



DB instance identifier
Type a name for your DB instance. The name must be unique across all DB instances owned by your AWS account in the current AWS Region.

database-1

The DB instance identifier is case-insensitive, but is stored as all lowercase (as in "mydbinstance"). Constraints: 1 to 60 alphanumeric characters or hyphens. First character must be a letter. Can't contain two consecutive hyphens. Can't end with a hyphen.

Master username Info
Type a login ID for the master user of your DB instance.

admin

1 to 16 alphanumeric characters. First character must be a letter.

□ Auto generate a password
Amazon RDS can generate a password for you, or you can specify your own password.

Master password Info

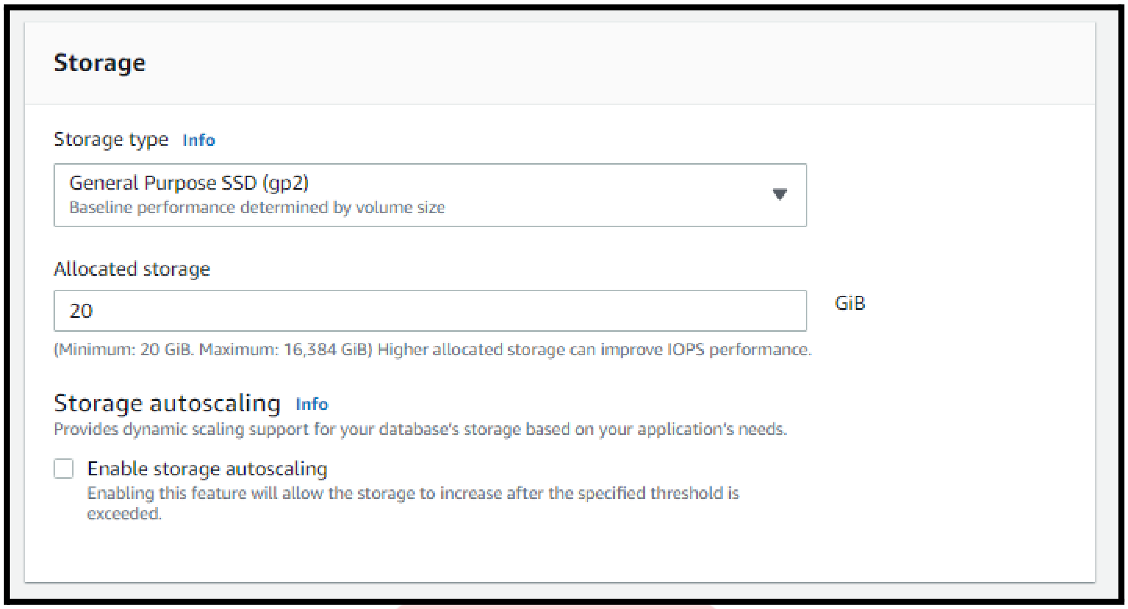
••••••••••

Constraints: At least 8 printable ASCII characters. Can't contain any of the following: / (slash), '(single quote), "(double quote) and @ (at sign).

Confirm password Info

••••••••••

In the storage section, we set the minimum allowed to 20 GiB and make sure that the "Enable storage autoscaling" box is unchecked.



Storage

Storage type Info

General Purpose SSD (gp2)
Baseline performance determined by volume size

Allocated storage

20 GiB

(Minimum: 20 GiB. Maximum: 16,384 GiB) Higher allocated storage can improve IOPS performance.

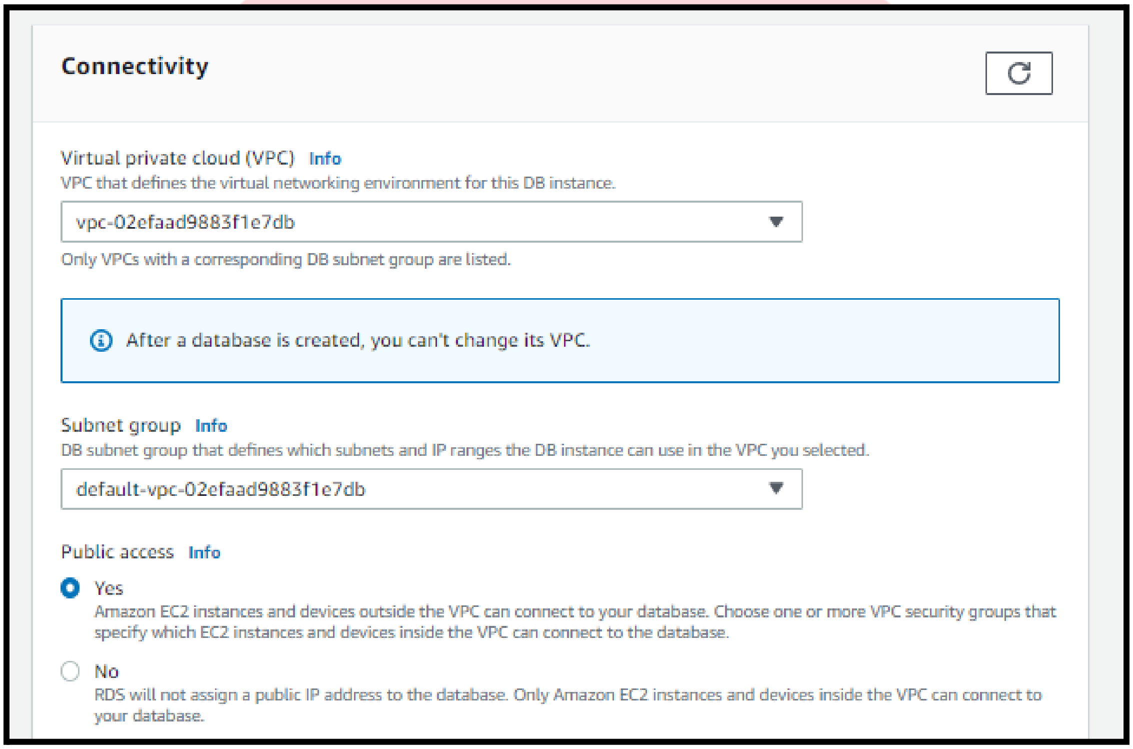
Storage autoscaling Info

Provides dynamic scaling support for your database's storage based on your application's needs.

□ Enable storage autoscaling
Enabling this feature will allow the storage to increase after the specified threshold is exceeded.

e

In the connectivity section, we use all default options and allow public access to make the process of connecting to the database easier.



Connectivity

Virtual private cloud (VPC) Info

VPC that defines the virtual networking environment for this DB instance.

vpc-02efaad9883f1e7db

Only VPCs with a corresponding DB subnet group are listed.

After a database is created, you can't change its VPC.

Subnet group Info

DB subnet group that defines which subnets and IP ranges the DB instance can use in the VPC you selected.

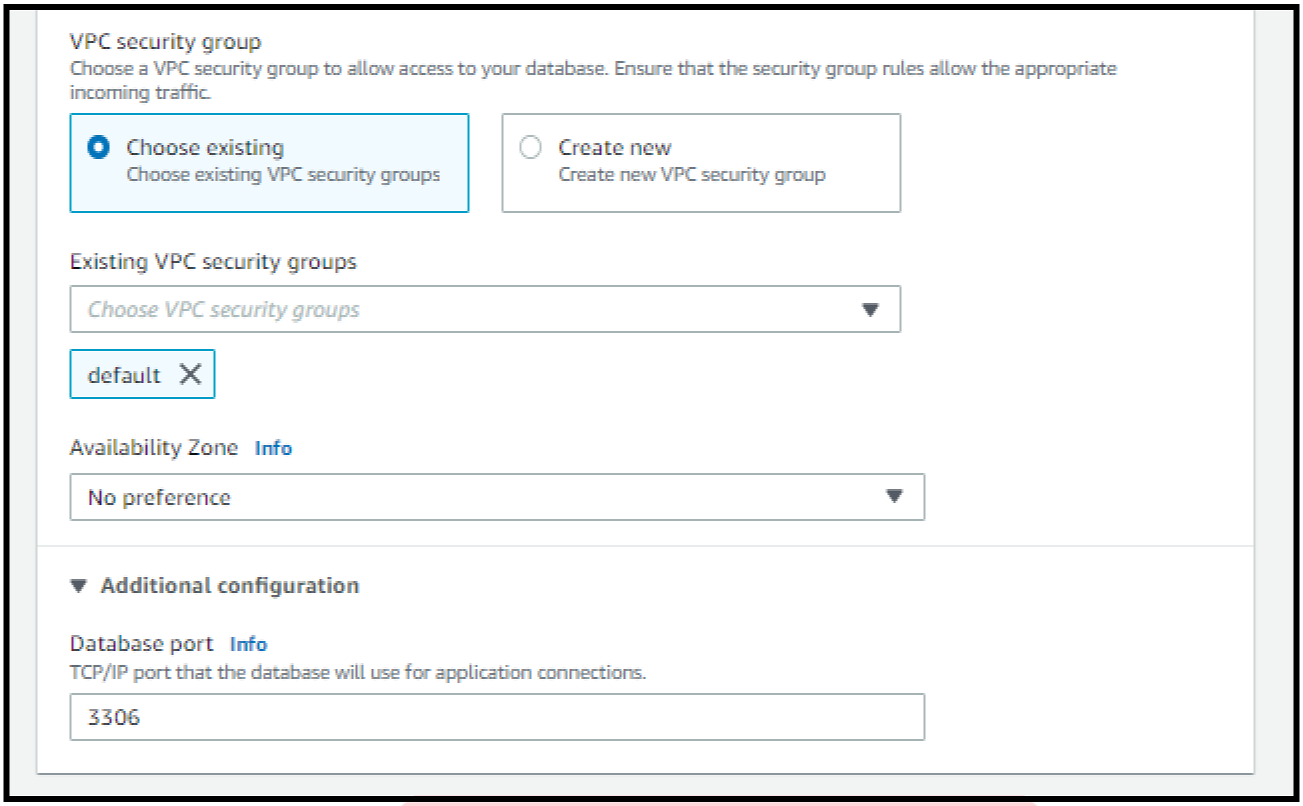
default-vpc-02efaad9883f1e7db

Public access Info

● Yes
Amazon EC2 instances and devices outside the VPC can connect to your database. Choose one or more VPC security groups that specify which EC2 instances and devices inside the VPC can connect to the database.

○ No
RDS will not assign a public IP address to the database. Only Amazon EC2 instances and devices inside the VPC can connect to your database.

We select the default VPC and leave everything else as it is.



VPC security group

Choose a VPC security group to allow access to your database. Ensure that the security group rules allow the appropriate incoming traffic.

● Choose existing
Choose existing VPC security groups

○ Create new
Create new VPC security group

Existing VPC security groups

Choose VPC security groups ▼

default ✕

Availability Zone Info

No preference ▼

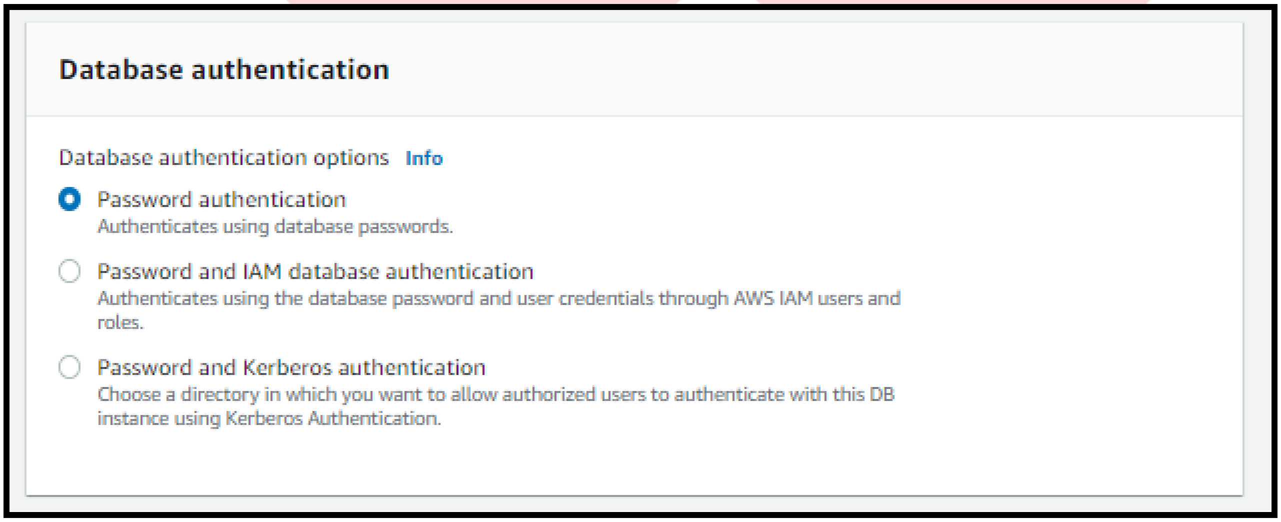
▼ Additional configuration

Database port Info

TCP/IP port that the database will use for application connections.

3306

We set the” Database authentication” with the password option. We leave everything else with the default configuration. We click on “create database” and wait for it to finish. Launching a database with our previously defined settings will take a few minutes.



Database authentication

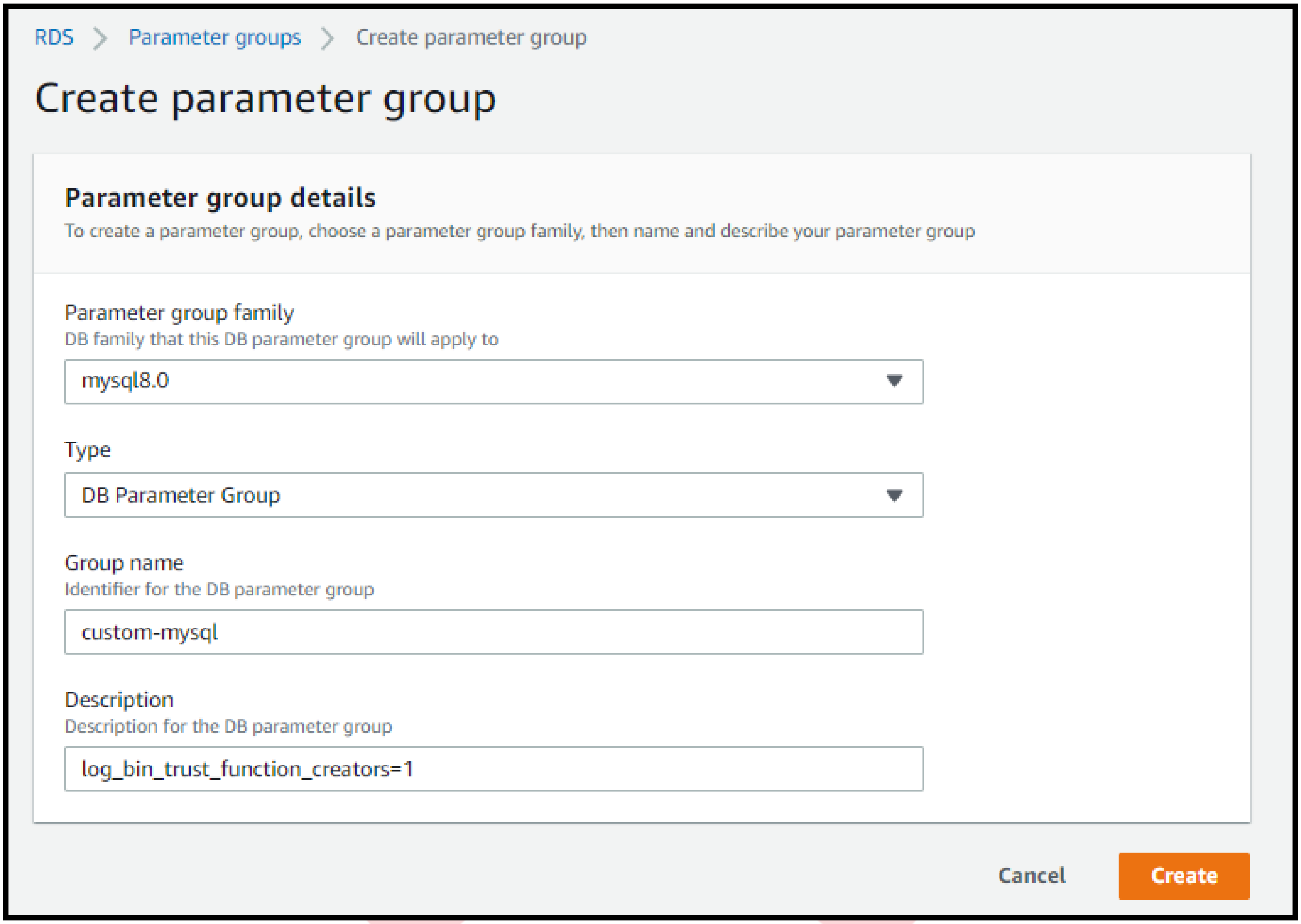
Database authentication options Info

● Password authentication
Authenticates using database passwords.

○ Password and IAM database authentication
Authenticates using the database password and user credentials through AWS IAM users and roles.

○ Password and Kerberos authentication
Choose a directory in which you want to allow authorized users to authenticate with this DB instance using Kerberos Authentication.

When the database is ready, we must create a new parameter group to run Sakila scripts successfully.



Create parameter group

Parameter group details

To create a parameter group, choose a parameter group family, then name and describe your parameter group

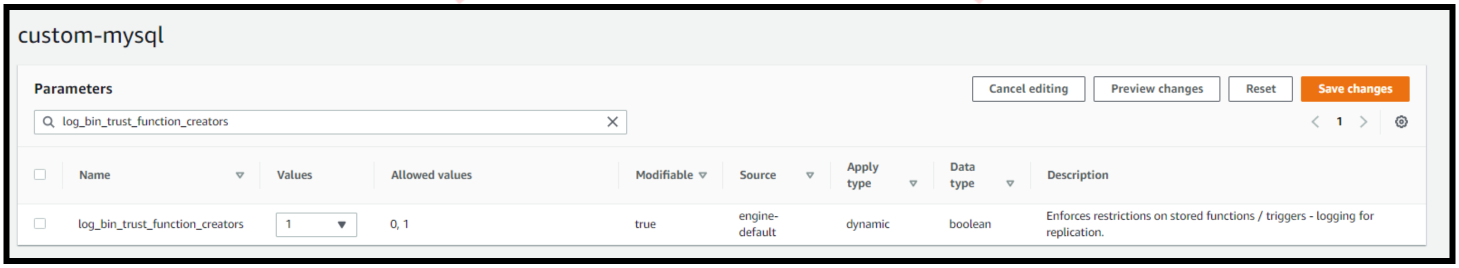
Parameter group family
DB family that this DB parameter group will apply to
mysql8.0

Type
DB Parameter Group

Group name
Identifier for the DB parameter group
custom-mysql

Description
Description for the DB parameter group
log_bin_trust_function_creators=1

In our new parameter group, we set **log_bin_trust_function_creators** to “1” and save the changes.



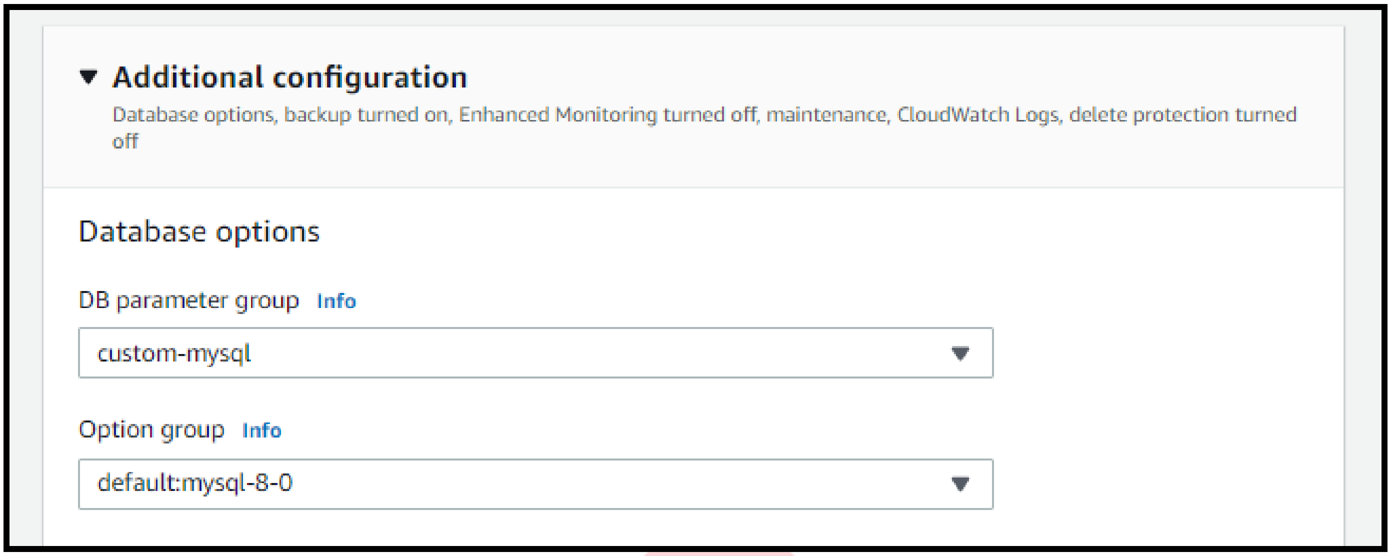
custom-mysql

Parameters

log_bin_trust_function_creators

□	Name	Values	Allowed values	Modifiable	Source	Apply type	Data type	Description
□	log_bin_trust_function_creators	1	0, 1	true	engine-default	dynamic	boolean	Enforces restrictions on stored functions / triggers - logging for replication.

Then we modify our existing database and set the parameter group to our custom one.



▼ **Additional configuration**

Database options, backup turned on, Enhanced Monitoring turned off, maintenance, CloudWatch Logs, delete protection turned off

Database options

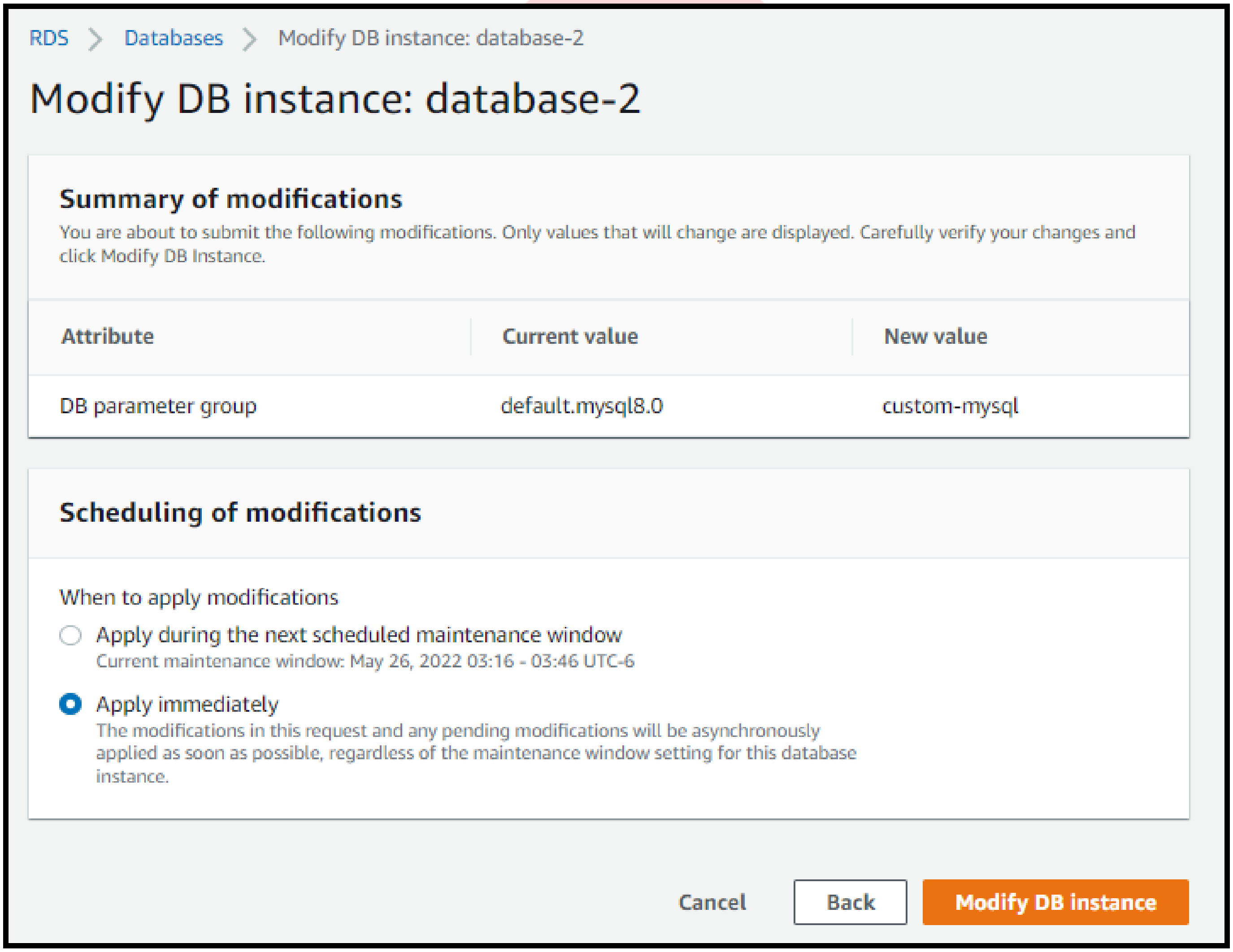
DB parameter group

custom-mysql ▼

Option group

default:mysql-8-0 ▼

We click on the "Modify DB" instance and check on "Apply immediately."



Modify DB instance: database-2

Summary of modifications

You are about to submit the following modifications. Only values that will change are displayed. Carefully verify your changes and click Modify DB Instance.

Attribute	Current value	New value
DB parameter group	default.mysql8.0	custom-mysql

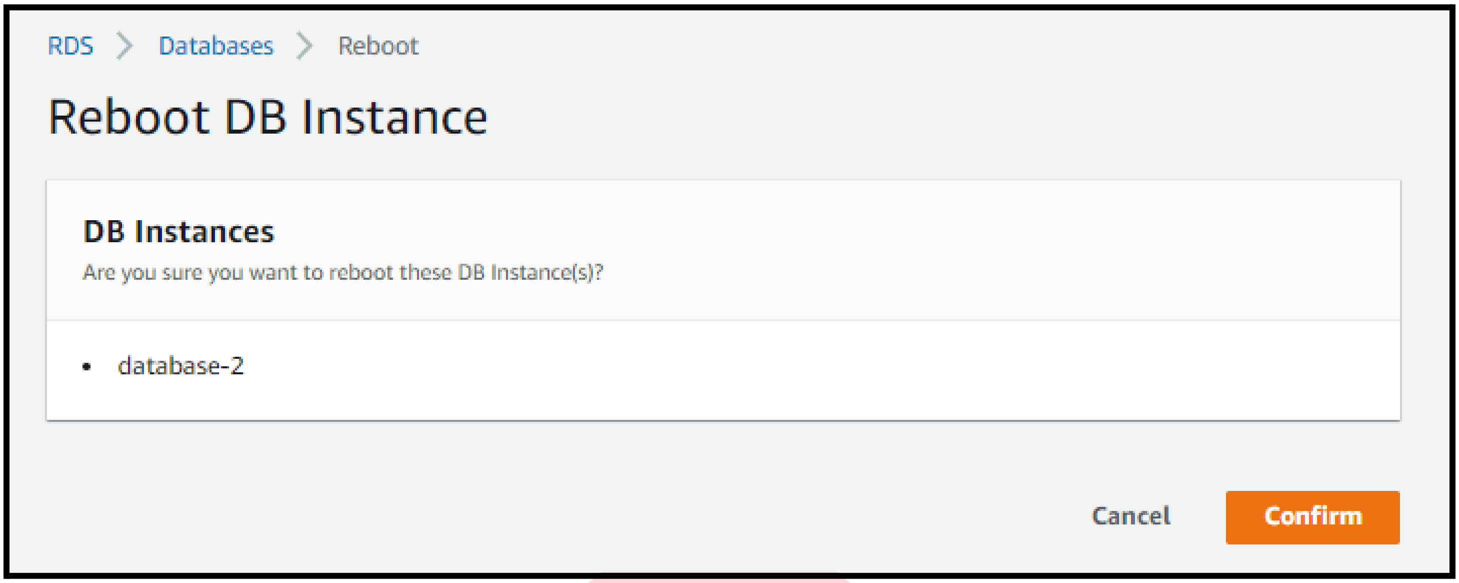
Scheduling of modifications

When to apply modifications

○ Apply during the next scheduled maintenance window
Current maintenance window: May 26, 2022 03:16 - 03:46 UTC-6

● Apply immediately
The modifications in this request and any pending modifications will be asynchronously applied as soon as possible, regardless of the maintenance window setting for this database instance.

Then we proceed to reboot our instance to apply our latest change.



RDS > Databases > Reboot

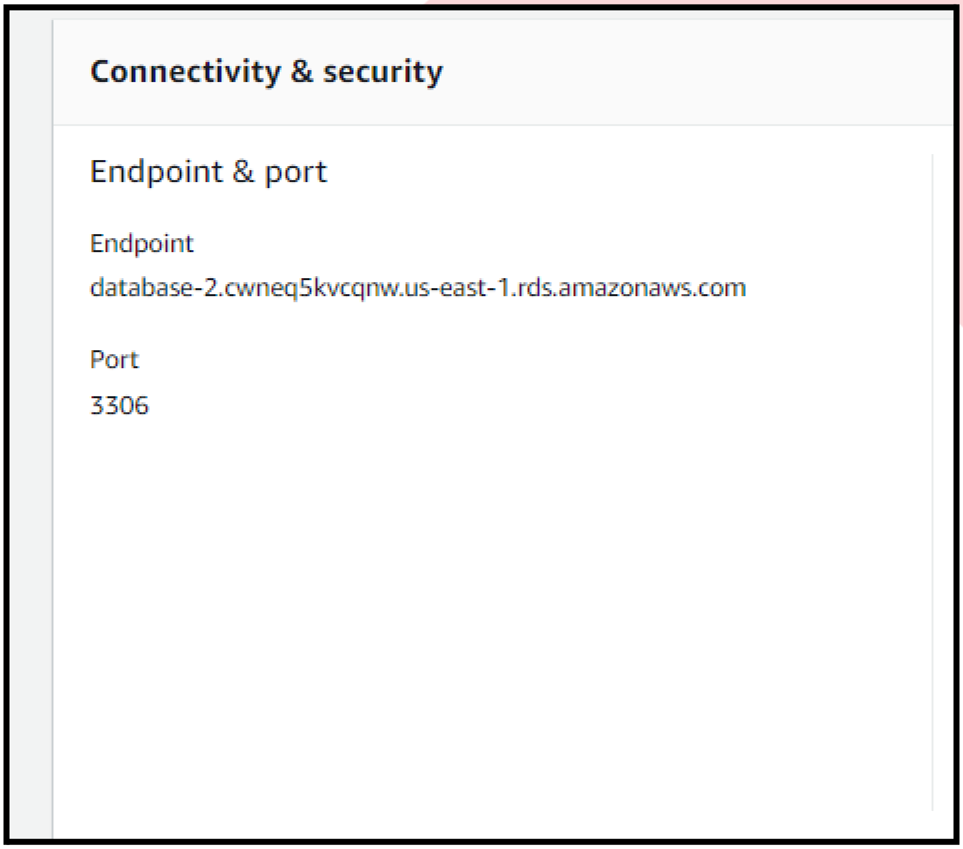
Reboot DB Instance

DB Instances

Are you sure you want to reboot these DB Instance(s)?

- database-2

Once the reboot is over, we click on our database to get more details and the endpoints to set the connection. (Don’t worry, this DB will no longer exist by the time you’re reading this.)



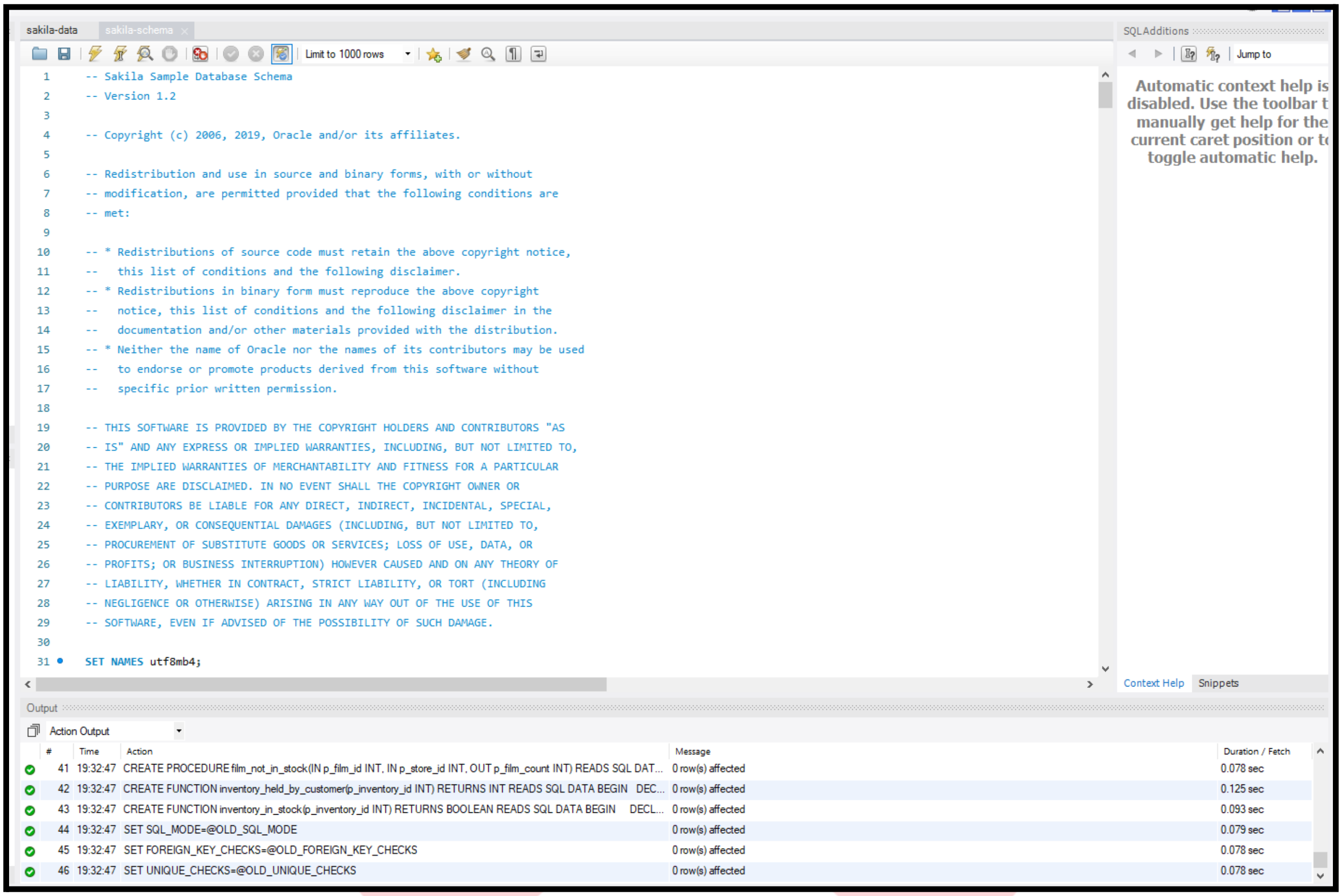
Connectivity & security

Endpoint & port

Endpoint
database-2.cwneq5kvcqnw.us-east-1.rds.amazonaws.com

Port
3306

With the information we got, we can establish a connection using MySQL Workbench, DBeaver, etc. We run the Sakila schema script first and then the Sakila data to finish setting up our data source.

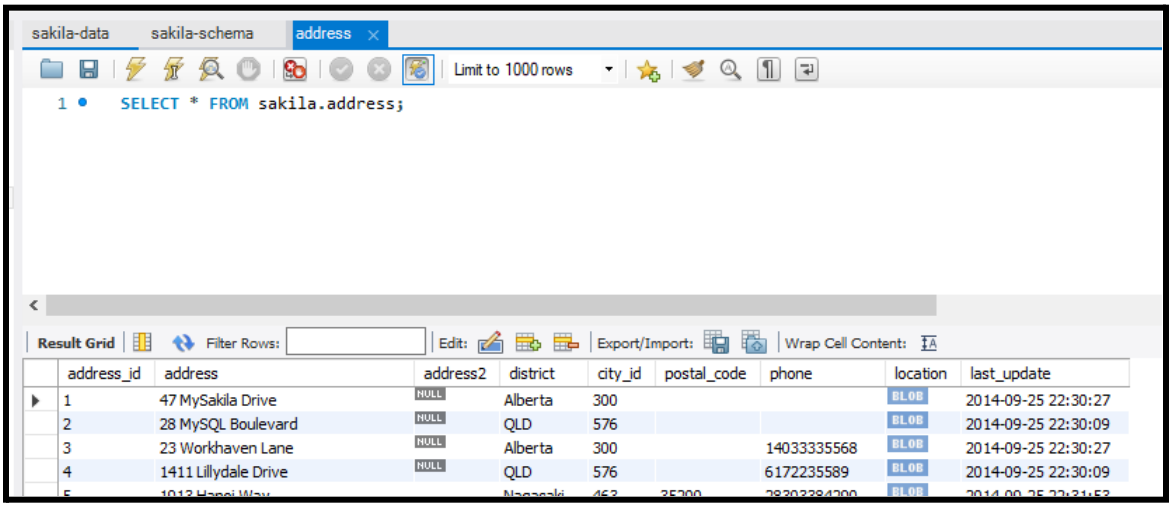


```
-- Sakila Sample Database Schema
-- Version 1.2

-- Copyright (c) 2006, 2019, Oracle and/or its affiliates.

-- Redistribution and use in source and binary forms, with or without
-- modification, are permitted provided that the following conditions are
-- met:

-- * Redistributions of source code must retain the above copyright notice,
--   this list of conditions and the following disclaimer.
-- * Redistributions in binary form must reproduce the above copyright
--   notice, this list of conditions and the following disclaimer in the
--   documentation and/or other materials provided with the distribution.
-- * Neither the name of Oracle nor the names of its contributors may be used
--   to endorse or promote products derived from this software without
--   specific prior written permission.

-- THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS
-- IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO,
-- THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR
-- PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR
-- CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL,
-- EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO,
-- PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR
-- PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF
-- LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING
-- NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS
-- SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

SET NAMES utf8mb4;
```

Automatic context help is disabled. Use the toolbar t manually get help for the current caret position or t toggle automatic help.

```
   #  Time      Action                                                                                           Message              Duration / Fetch
✓ 41  19:32:47  CREATE PROCEDURE film_not_in_stock(IN p_film_id INT, IN p_store_id INT, OUT p_film_count INT) READS SQL DAT...  0 row(s) affected  0.078 sec
✓ 42  19:32:47  CREATE FUNCTION inventory_held_by_customer(p_inventory_id INT) RETURNS INT READS SQL DATA BEGIN  DEC...  0 row(s) affected  0.125 sec
✓ 43  19:32:47  CREATE FUNCTION inventory_in_stock(p_inventory_id INT) RETURNS BOOLEAN READS SQL DATA BEGIN   DECL...  0 row(s) affected  0.093 sec
✓ 44  19:32:47  SET SQL_MODE=@OLD_SQL_MODE                                                                       0 row(s) affected  0.079 sec
✓ 45  19:32:47  SET FOREIGN_KEY_CHECKS=@OLD_FOREIGN_KEY_CHECKS                                                   0 row(s) affected  0.078 sec
✓ 46  19:32:47  SET UNIQUE_CHECKS=@OLD_UNIQUE_CHECKS                                                             0 row(s) affected  0.078 sec
```

We will have a database full of data, which we can use for our data lake.

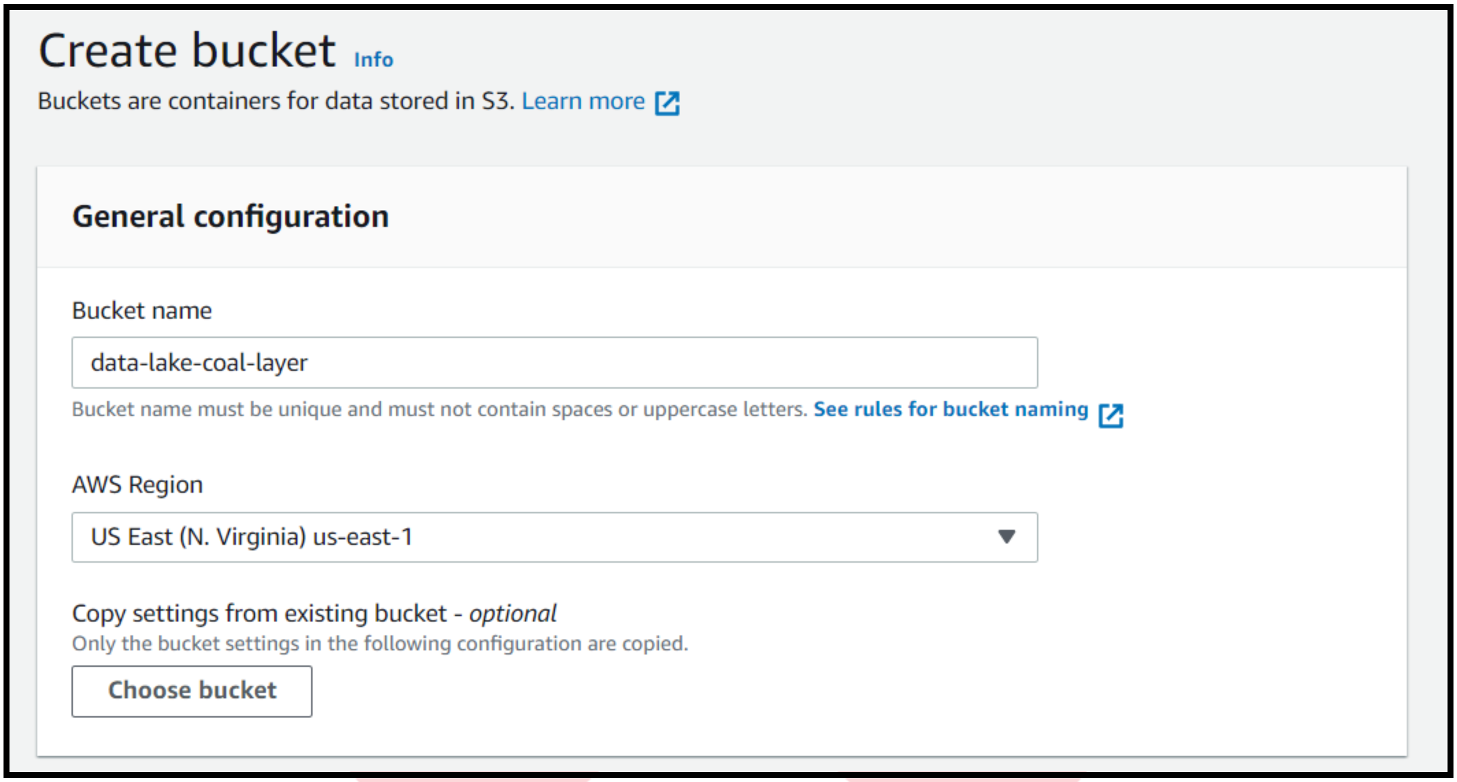


```
SELECT * FROM sakila.address;
```

address_id	address	address2	district	city_id	postal_code	phone	location	last_update
1	47 MySakila Drive	NULL	Alberta	300			BLOB	2014-09-25 22:30:27
2	28 MySQL Boulevard	NULL	QLD	576			BLOB	2014-09-25 22:30:09
3	23 Workhaven Lane	NULL	Alberta	300		14033335568	BLOB	2014-09-25 22:30:27
4	1411 Lillydale Drive	NULL	QLD	576		6172235589	BLOB	2014-09-25 22:30:09
5	1913 Hanoi Way		Nagasaki	463	35200	28303384290	BLOB	2014-09-25 22:31:53

2. Data Lake layers - S3

With the setup for our data source done, we can proceed to work on the layers. We will classify and store the data. We need to create three different layers for each stage of data.



Create bucket Info

Buckets are containers for data stored in S3. Learn more

General configuration

Bucket name

data-lake-coal-layer

Bucket name must be unique and must not contain spaces or uppercase letters. **See rules for bucket naming**

AWS Region

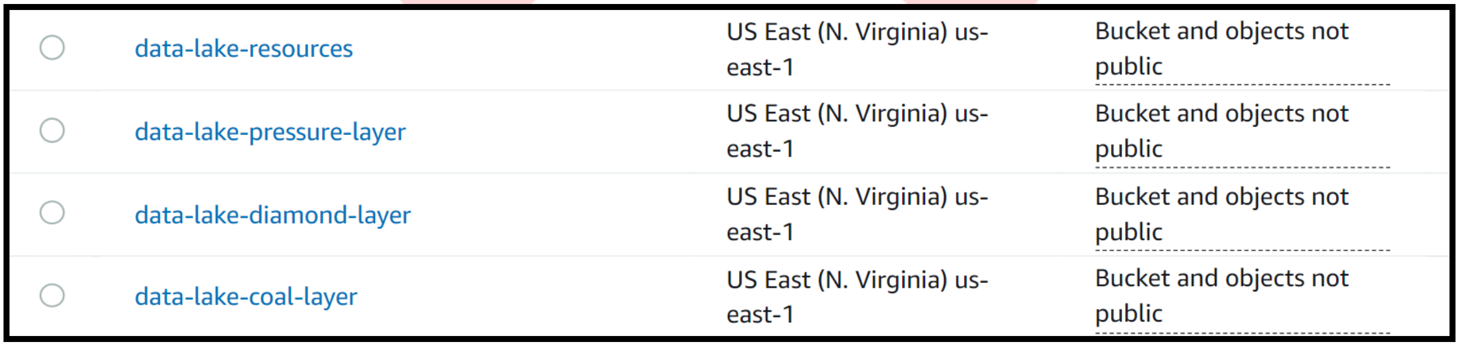
US East (N. Virginia) us-east-1 ▼

Copy settings from existing bucket - *optional*

Only the bucket settings in the following configuration are copied.

Choose bucket

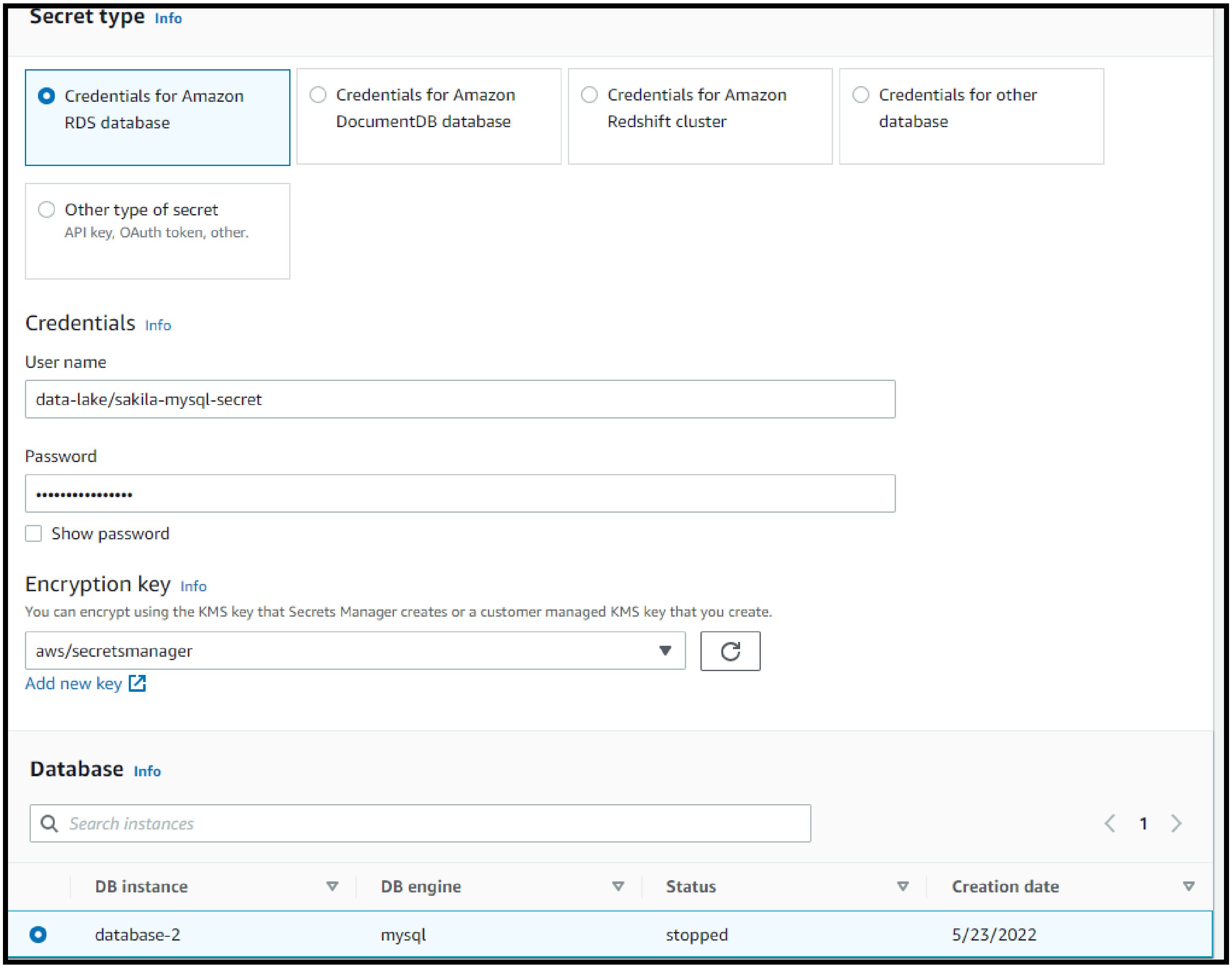
We create another bucket for the resources we will need in the future. We will end up with a setup like this.



○	data-lake-resources	US East (N. Virginia) us-east-1	Bucket and objects not public
○	data-lake-pressure-layer	US East (N. Virginia) us-east-1	Bucket and objects not public
○	data-lake-diamond-layer	US East (N. Virginia) us-east-1	Bucket and objects not public
○	data-lake-coal-layer	US East (N. Virginia) us-east-1	Bucket and objects not public

3. Credentials - Secret Manager

We set the name of our secret, and if our database is in RDS, we only need to select it and add the password. We can also store credentials for other databases.



Secret type Info

● Credentials for Amazon RDS database

○ Credentials for Amazon DocumentDB database

○ Credentials for Amazon Redshift cluster

○ Credentials for other database

○ Other type of secret
API key, OAuth token, other.

Credentials Info

User name: data-lake/sakila-mysql-secret

Password: ••••••••••••••••

□ Show password

Encryption key Info

You can encrypt using the KMS key that Secrets Manager creates or a customer managed KMS key that you create.

aws/secretsmanager

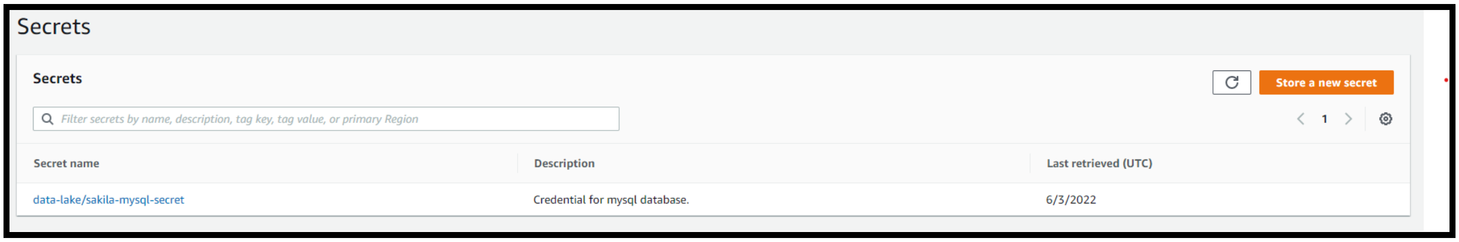
Add new key

Database Info

Search instances

	DB instance	DB engine	Status	Creation date
●	database-2	mysql	stopped	5/23/2022

In the end, our secret will look like this.



Secrets

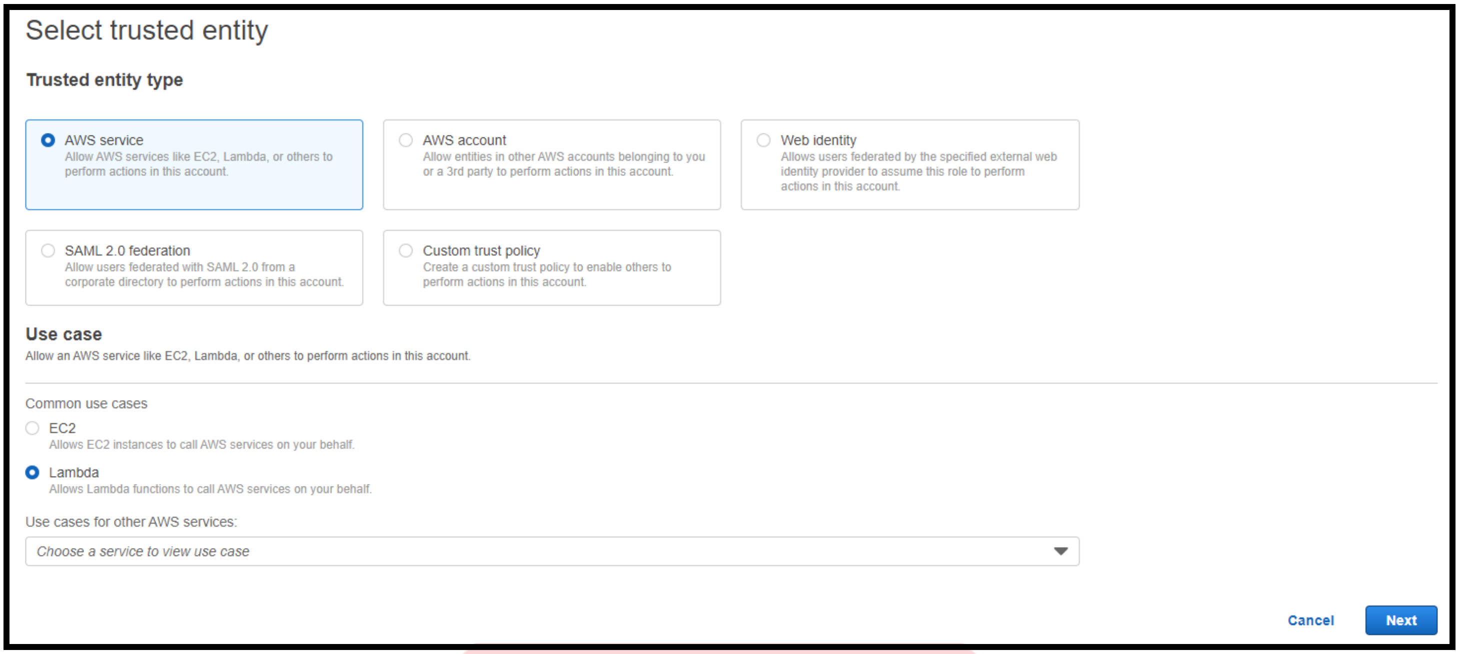
Secrets

Filter secrets by name, description, tag key, tag value, or primary Region

Secret name	Description	Last retrieved (UTC)
data-lake/sakila-mysql-secret	Credential for mysql database.	6/3/2022

4. Permissions - IAM

We set up the required IAM role for our lambda functions.



Select trusted entity

Trusted entity type

- ● AWS service — Allow AWS services like EC2, Lambda, or others to perform actions in this account.
- ○ AWS account — Allow entities in other AWS accounts belonging to you or a 3rd party to perform actions in this account.
- ○ Web identity — Allows users federated by the specified external web identity provider to assume this role to perform actions in this account.
- ○ SAML 2.0 federation — Allow users federated with SAML 2.0 from a corporate directory to perform actions in this account.
- ○ Custom trust policy — Create a custom trust policy to enable others to perform actions in this account.

Use case

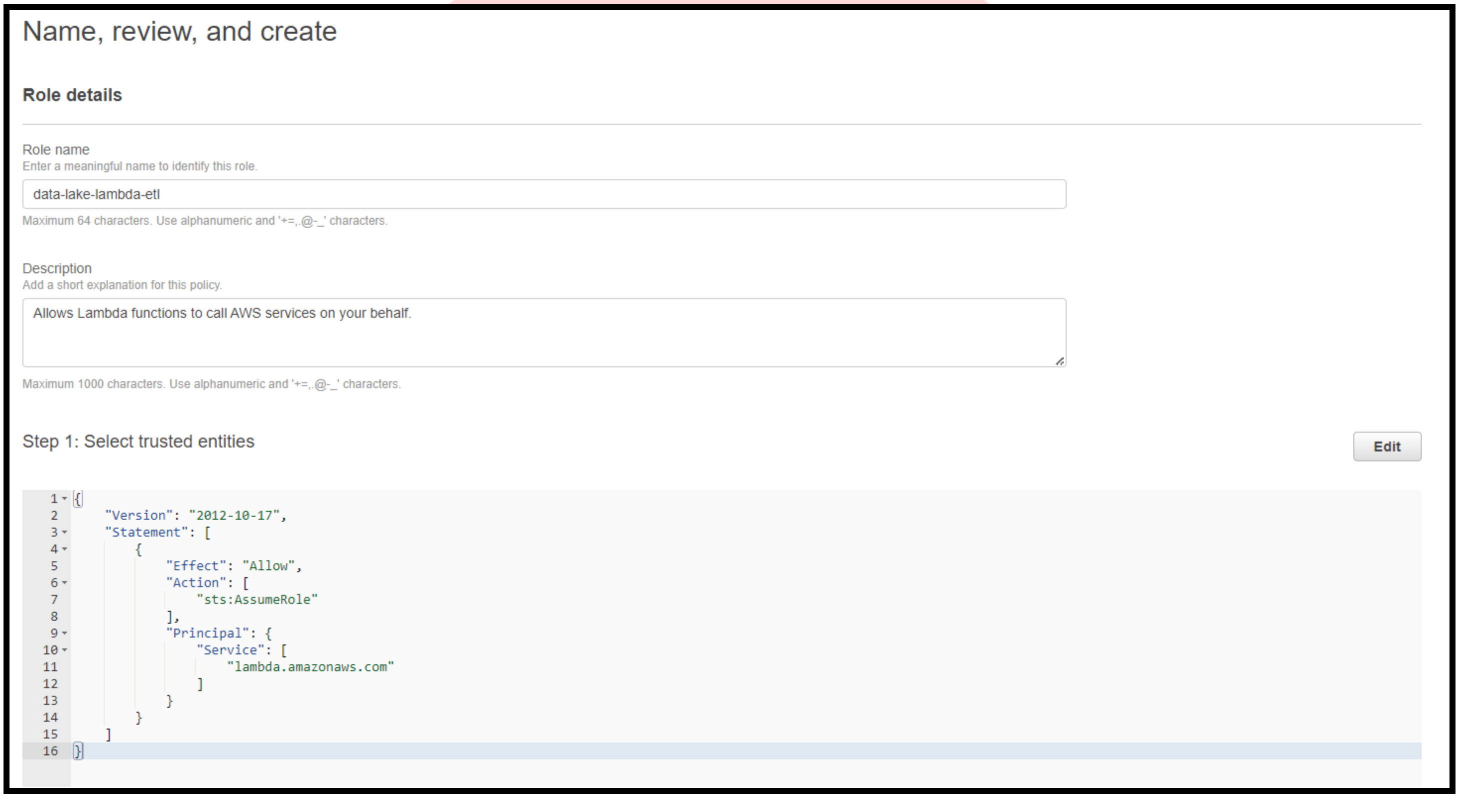
Allow an AWS service like EC2, Lambda, or others to perform actions in this account.

Common use cases

- ○ EC2 — Allows EC2 instances to call AWS services on your behalf.
- ● Lambda — Allows Lambda functions to call AWS services on your behalf.

Use cases for other AWS services:

Choose a service to view use case ▼



Name, review, and create

Role details

Role name

Enter a meaningful name to identify this role.

data-lake-lambda-etl

Maximum 64 characters. Use alphanumeric and '+=,.@-_' characters.

Description

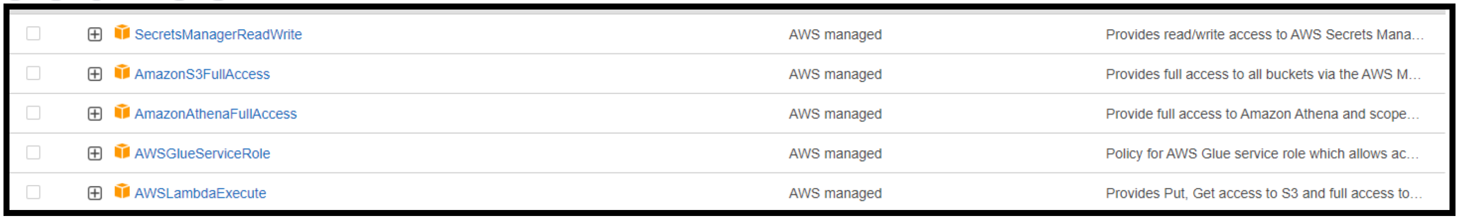
Add a short explanation for this policy.

Allows Lambda functions to call AWS services on your behalf.

Maximum 1000 characters. Use alphanumeric and '+=,.@-_' characters.

Step 1: Select trusted entities

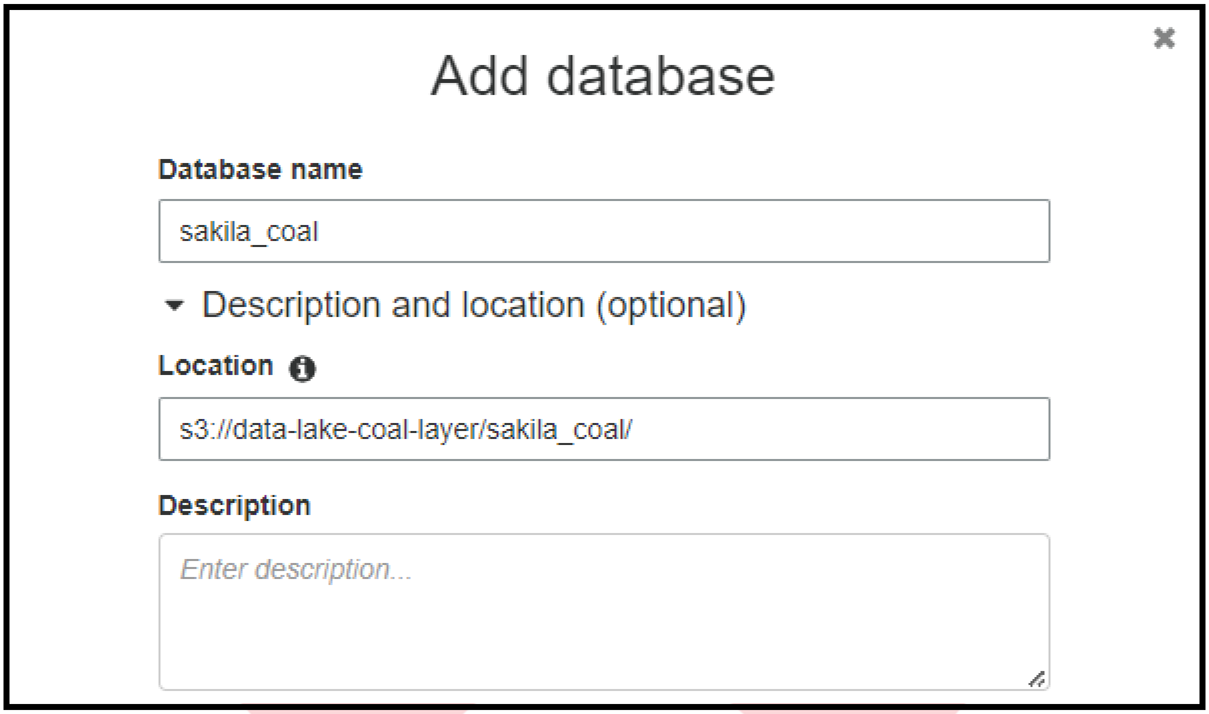
```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "sts:AssumeRole"
            ],
            "Principal": {
                "Service": [
                    "lambda.amazonaws.com"
                ]
            }
        }
    ]
}
```



□	⊞ SecretsManagerReadWrite	AWS managed	Provides read/write access to AWS Secrets Mana...
□	⊞ AmazonS3FullAccess	AWS managed	Provides full access to all buckets via the AWS M...
□	⊞ AmazonAthenaFullAccess	AWS managed	Provide full access to Amazon Athena and scope...
□	⊞ AWSGlueServiceRole	AWS managed	Policy for AWS Glue service role which allows ac...
□	⊞ AWSLambdaExecute	AWS managed	Provides Put, Get access to S3 and full access to...

5. Setting up Glue

We create a database for our data source and each layer in the data lake.



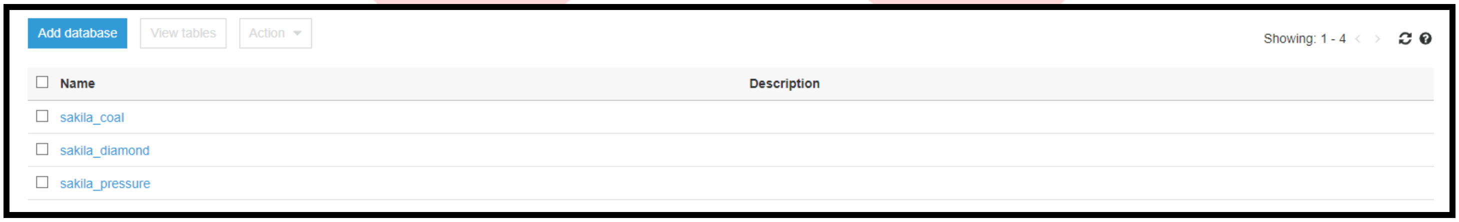
Add database

Database name: sakila_coal

Description and location (optional)

Location: s3://data-lake-coal-layer/sakila_coal/

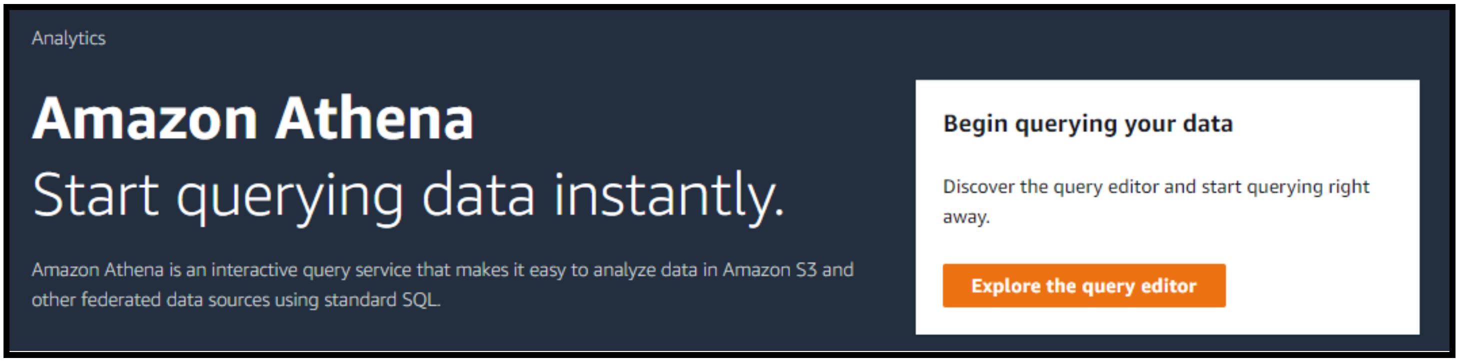
Description: Enter description...



□ Name	Description
□ sakila_coal	
□ sakila_diamond	
□ sakila_pressure	

6. Querying Data – Amazon Athena

If it is the first time we set up Athena, we need to finish the initial configuration.



Analytics

Amazon Athena

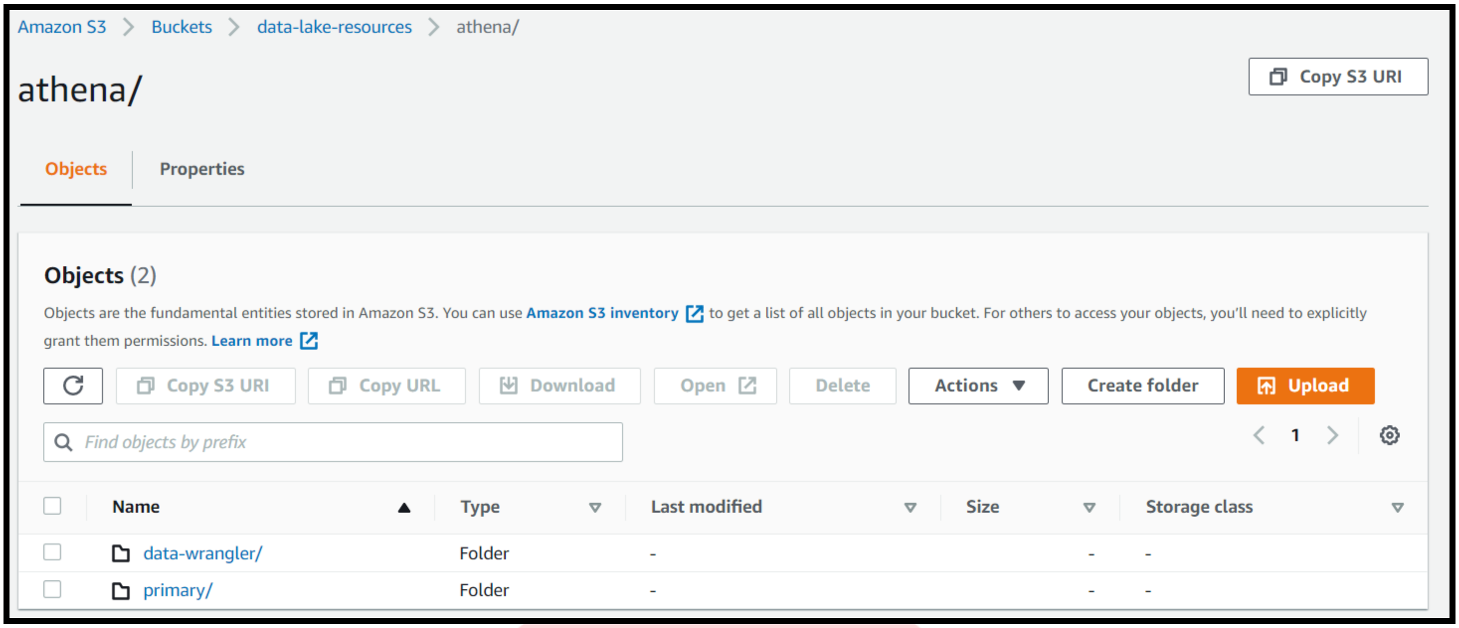
Start querying data instantly.

Amazon Athena is an interactive query service that makes it easy to analyze data in Amazon S3 and other federated data sources using standard SQL.

Begin querying your data

Discover the query editor and start querying right away.

We need to set up a location for the query results to be stored on. We will use this one:



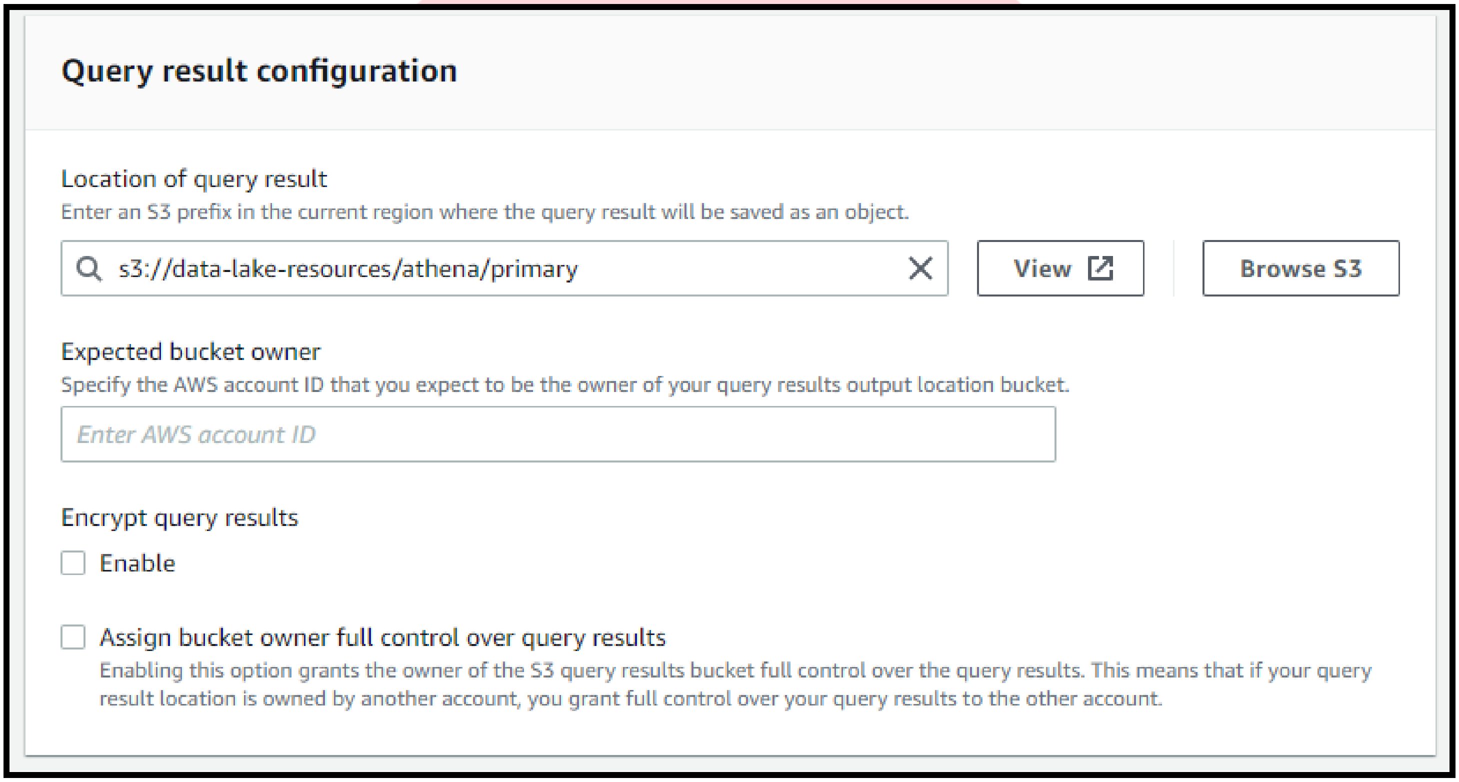
athena/

Objects (2)

Objects are the fundamental entities stored in Amazon S3. You can use **Amazon S3 inventory** to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. **Learn more**

□	Name ▲	Type	Last modified	Size	Storage class
□	data-wrangler/	Folder	-	-	-
□	primary/	Folder	-	-	-

We set the path for our primary workspace.



Query result configuration

Location of query result

Enter an S3 prefix in the current region where the query result will be saved as an object.

s3://data-lake-resources/athena/primary

Expected bucket owner

Specify the AWS account ID that you expect to be the owner of your query results output location bucket.

Enter AWS account ID

Encrypt query results

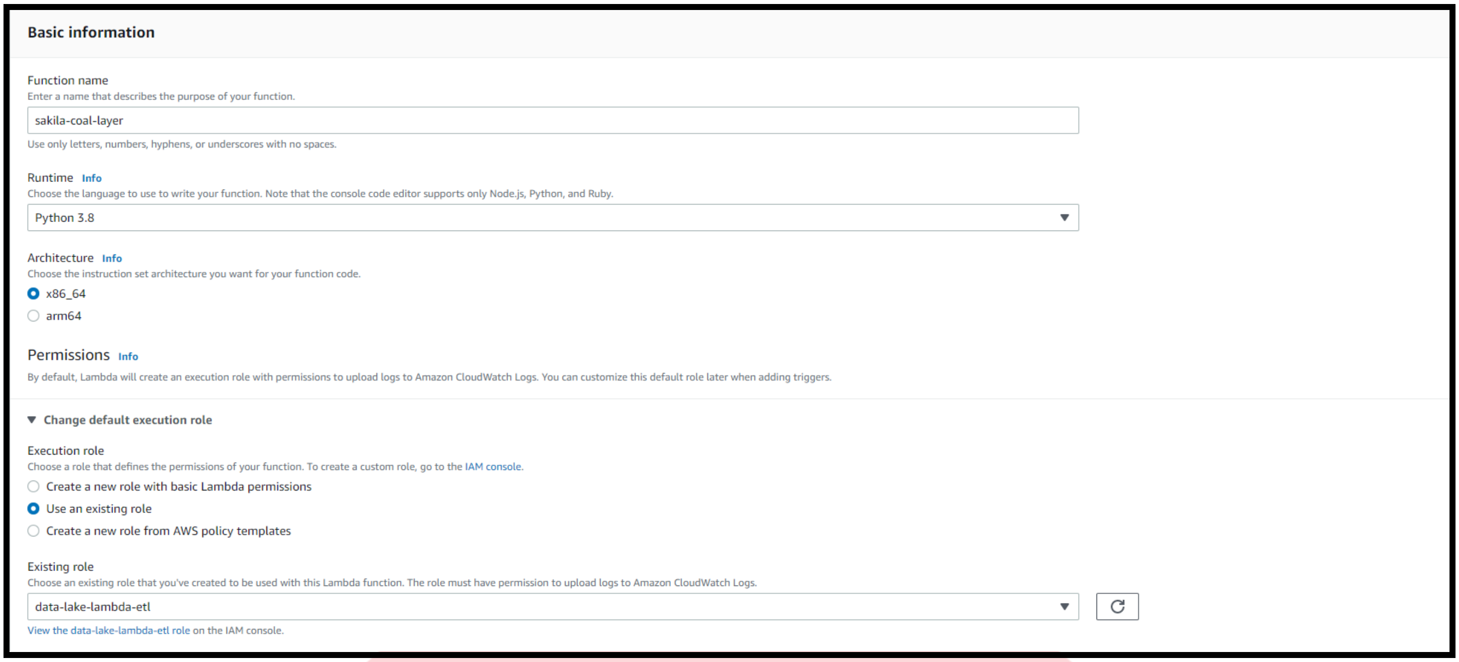
□ Enable

□ Assign bucket owner full control over query results

Enabling this option grants the owner of the S3 query results bucket full control over the query results. This means that if your query result location is owned by another account, you grant full control over your query results to the other account.

7. Data Processing - Lambda

Then we create a new lambda function for each layer. We use the settings specified in the article.



Basic information

Function name
Enter a name that describes the purpose of your function.
sakila-coal-layer
Use only letters, numbers, hyphens, or underscores with no spaces.

Runtime Info
Choose the language to use to write your function. Note that the console code editor supports only Node.js, Python, and Ruby.
Python 3.8 ▼

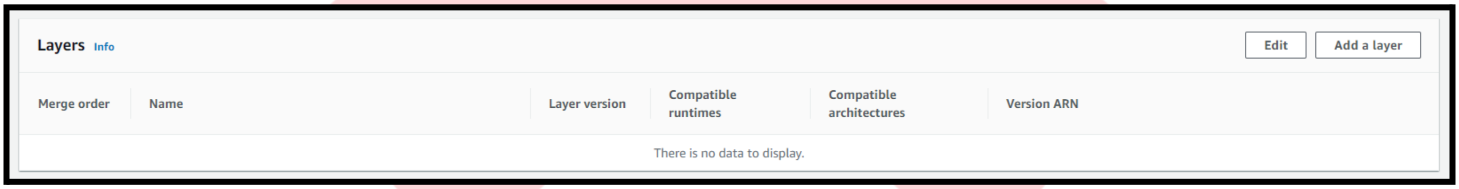
Architecture Info
Choose the instruction set architecture you want for your function code.
● x86_64
○ arm64

Permissions Info
By default, Lambda will create an execution role with permissions to upload logs to Amazon CloudWatch Logs. You can customize this default role later when adding triggers.

▼ **Change default execution role**

Execution role
Choose a role that defines the permissions of your function. To create a custom role, go to the IAM console.
○ Create a new role with basic Lambda permissions
● Use an existing role
○ Create a new role from AWS policy templates

Existing role
Choose an existing role that you've created to be used with this Lambda function. The role must have permission to upload logs to Amazon CloudWatch Logs.
data-lake-lambda-etl ▼
View the data-lake-lambda-etl role on the IAM console.

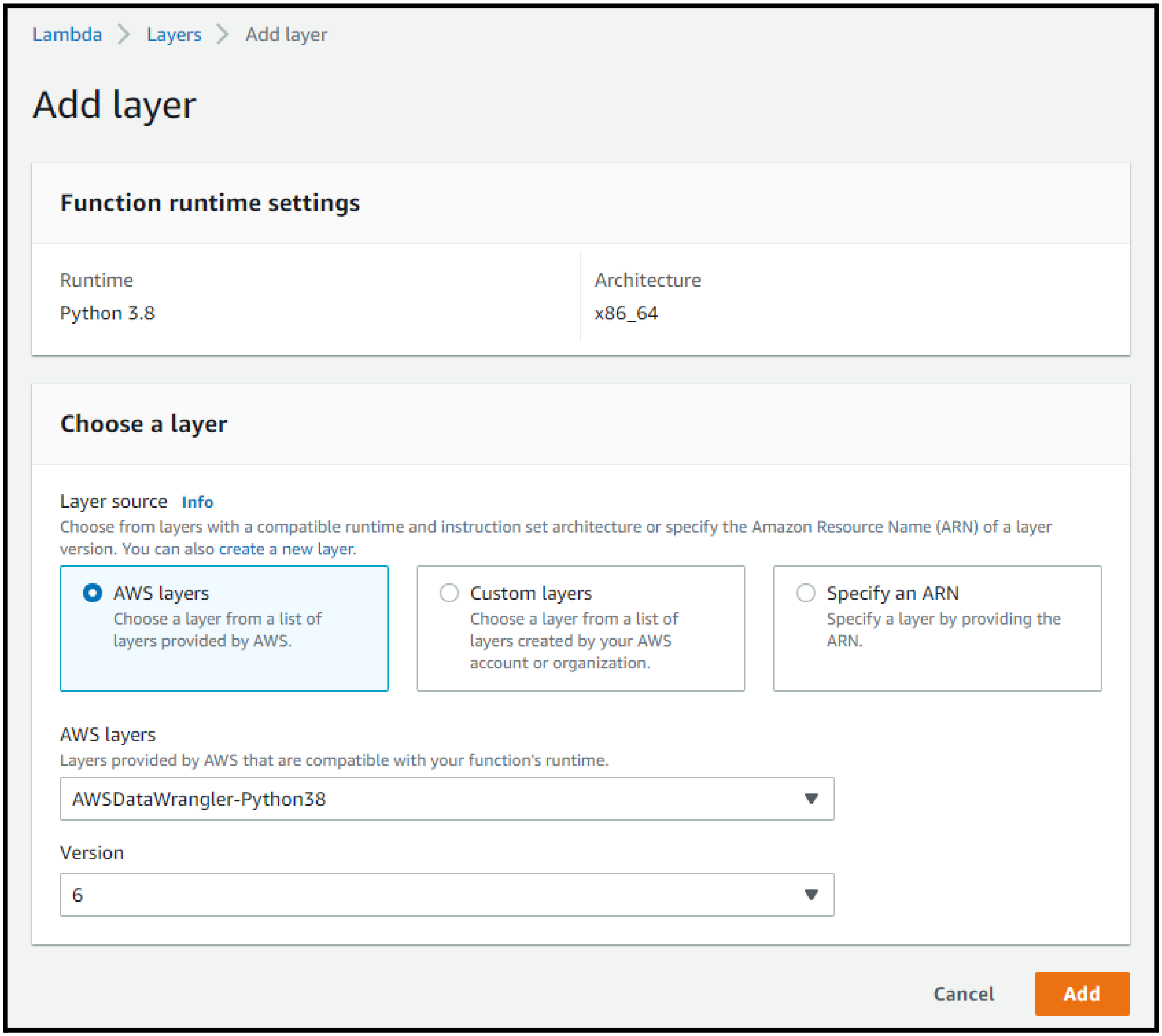


Layers Info

Merge order	Name	Layer version	Compatible runtimes	Compatible architectures	Version ARN

There is no data to display.

It is vital to add AWS Data wrangler with a layer provided by AWS.



Add layer

Function runtime settings

Runtime

Python 3.8

Architecture

x86_64

Choose a layer

Layer source

Choose from layers with a compatible runtime and instruction set architecture or specify the Amazon Resource Name (ARN) of a layer version. You can also create a new layer.

● AWS layers
Choose a layer from a list of layers provided by AWS.

○ Custom layers
Choose a layer from a list of layers created by your AWS account or organization.

○ Specify an ARN
Specify a layer by providing the ARN.

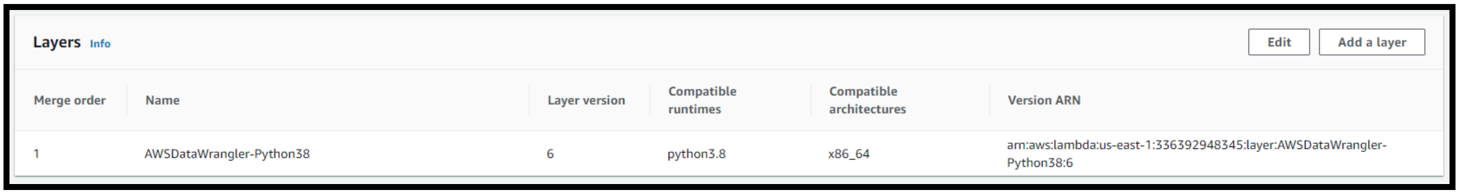
AWS layers

Layers provided by AWS that are compatible with your function's runtime.

AWSDataWrangler-Python38 ▼

Version

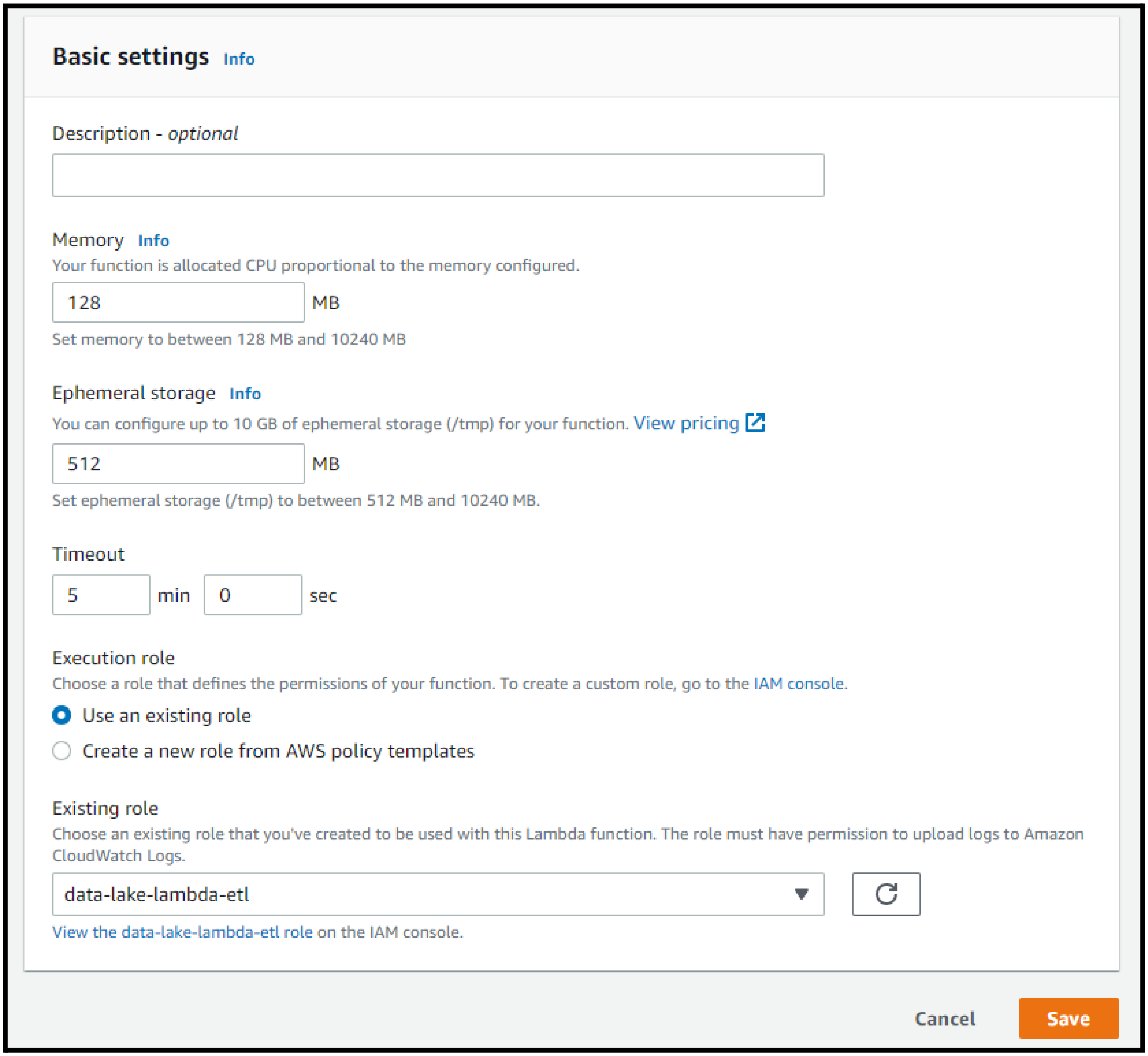
6 ▼



Layers

Merge order	Name	Layer version	Compatible runtimes	Compatible architectures	Version ARN
1	AWSDataWrangler-Python38	6	python3.8	x86_64	arn:aws:lambda:us-east-1:336392948345:layer:AWSDataWrangler-Python38:6

Finally, we set a good amount of memory and maximum execution time so we can run our lambda functions. Remember to add the code from the repository and update it with the values from your data lake.



Basic settings Info

Description - *optional*

Memory Info
Your function is allocated CPU proportional to the memory configured.
128 MB
Set memory to between 128 MB and 10240 MB

Ephemeral storage Info
You can configure up to 10 GB of ephemeral storage (/tmp) for your function. View pricing
512 MB
Set ephemeral storage (/tmp) to between 512 MB and 10240 MB.

Timeout
5 min 0 sec

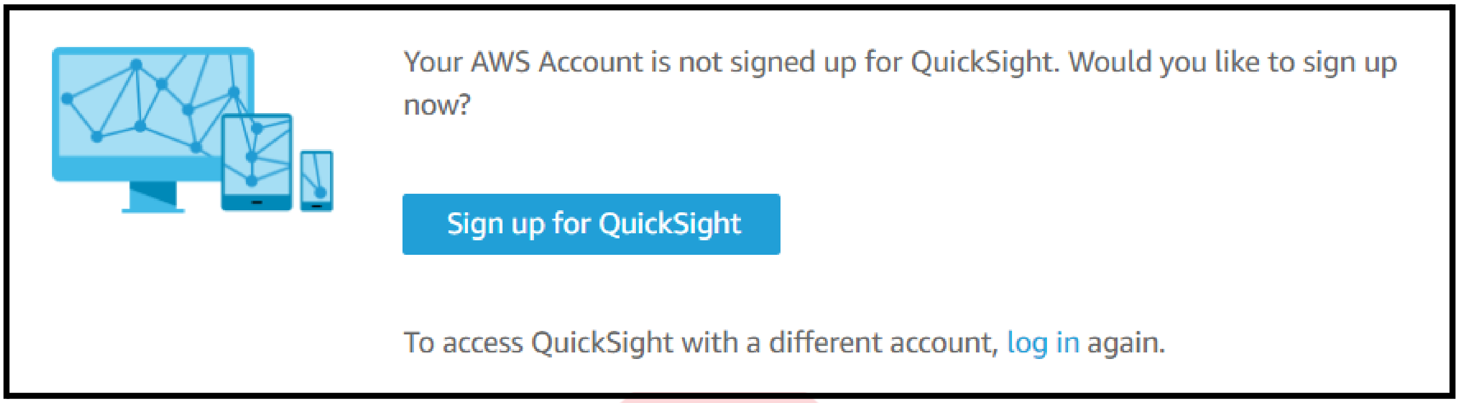
Execution role
Choose a role that defines the permissions of your function. To create a custom role, go to the IAM console.

● Use an existing role
○ Create a new role from AWS policy templates

Existing role
Choose an existing role that you've created to be used with this Lambda function. The role must have permission to upload logs to Amazon CloudWatch Logs.
data-lake-lambda-etl
View the data-lake-lambda-etl role on the IAM console.

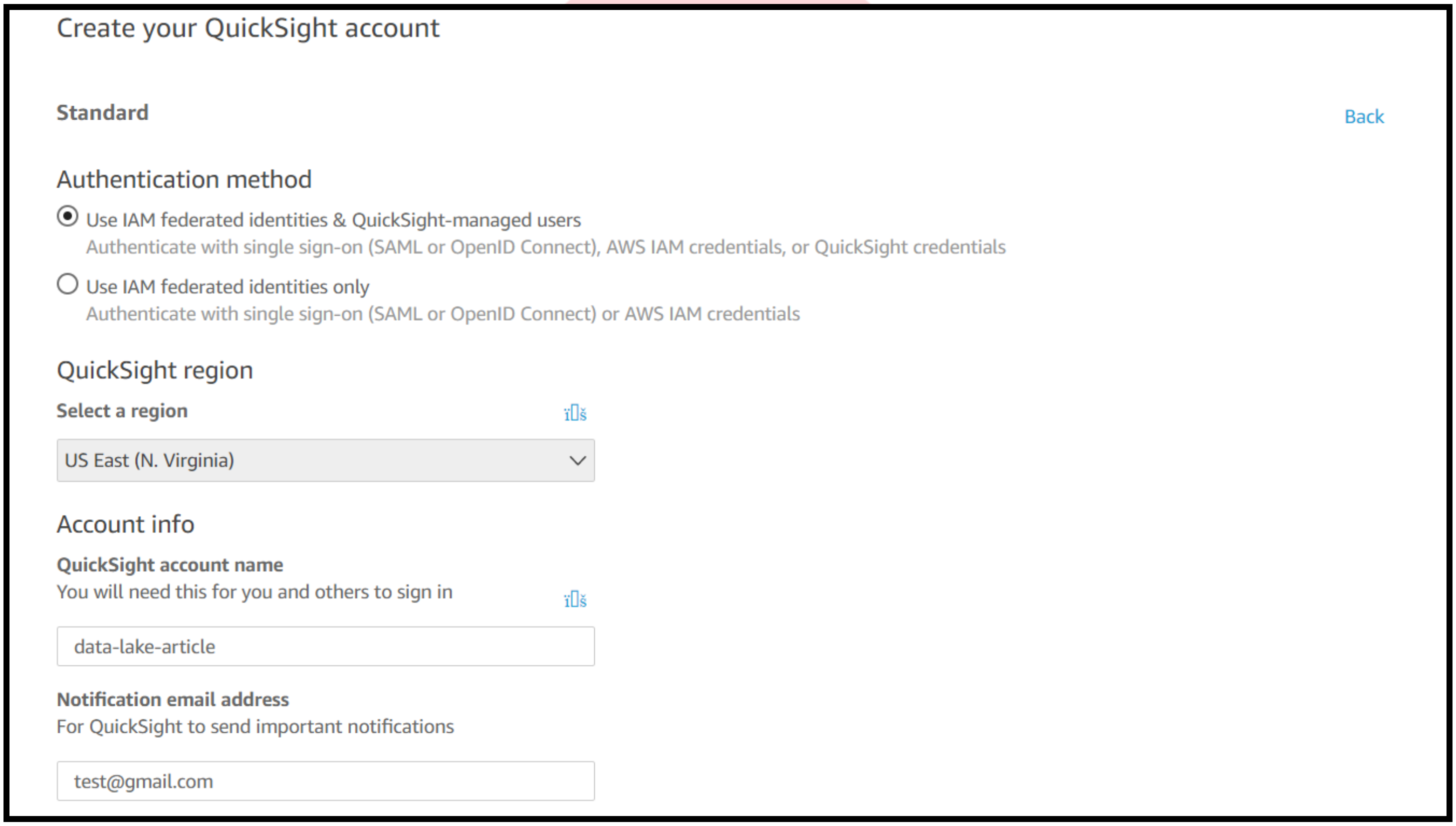
8. Business Intelligence - Quicksight

We need to sign up for Quicksight. We will make use of the 30-day trial for this project and set the following values;



Your AWS Account is not signed up for QuickSight. Would you like to sign up now?

To access QuickSight with a different account, log in again.



Create your QuickSight account

Standard

Authentication method

● Use IAM federated identities & QuickSight-managed users
Authenticate with single sign-on (SAML or OpenID Connect), AWS IAM credentials, or QuickSight credentials

○ Use IAM federated identities only
Authenticate with single sign-on (SAML or OpenID Connect) or AWS IAM credentials

QuickSight region

Select a region: US East (N. Virginia)

Account info

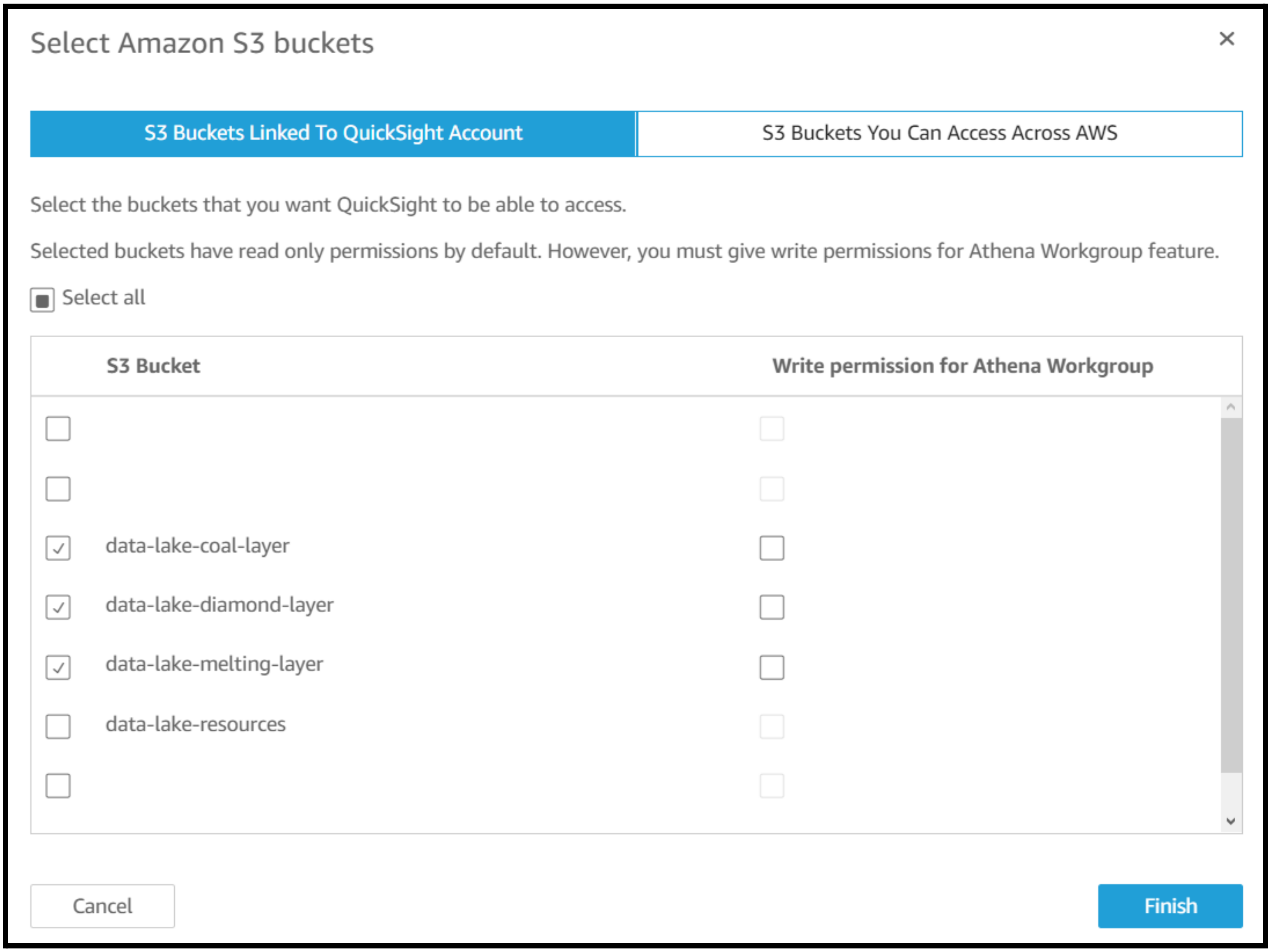
QuickSight account name
You will need this for you and others to sign in

data-lake-article

Notification email address
For QuickSight to send important notifications

test@gmail.com

Use your own buckets for this step.



Select Amazon S3 buckets

×
S3 Buckets Linked To QuickSight Account
S3 Buckets You Can Access Across AWS

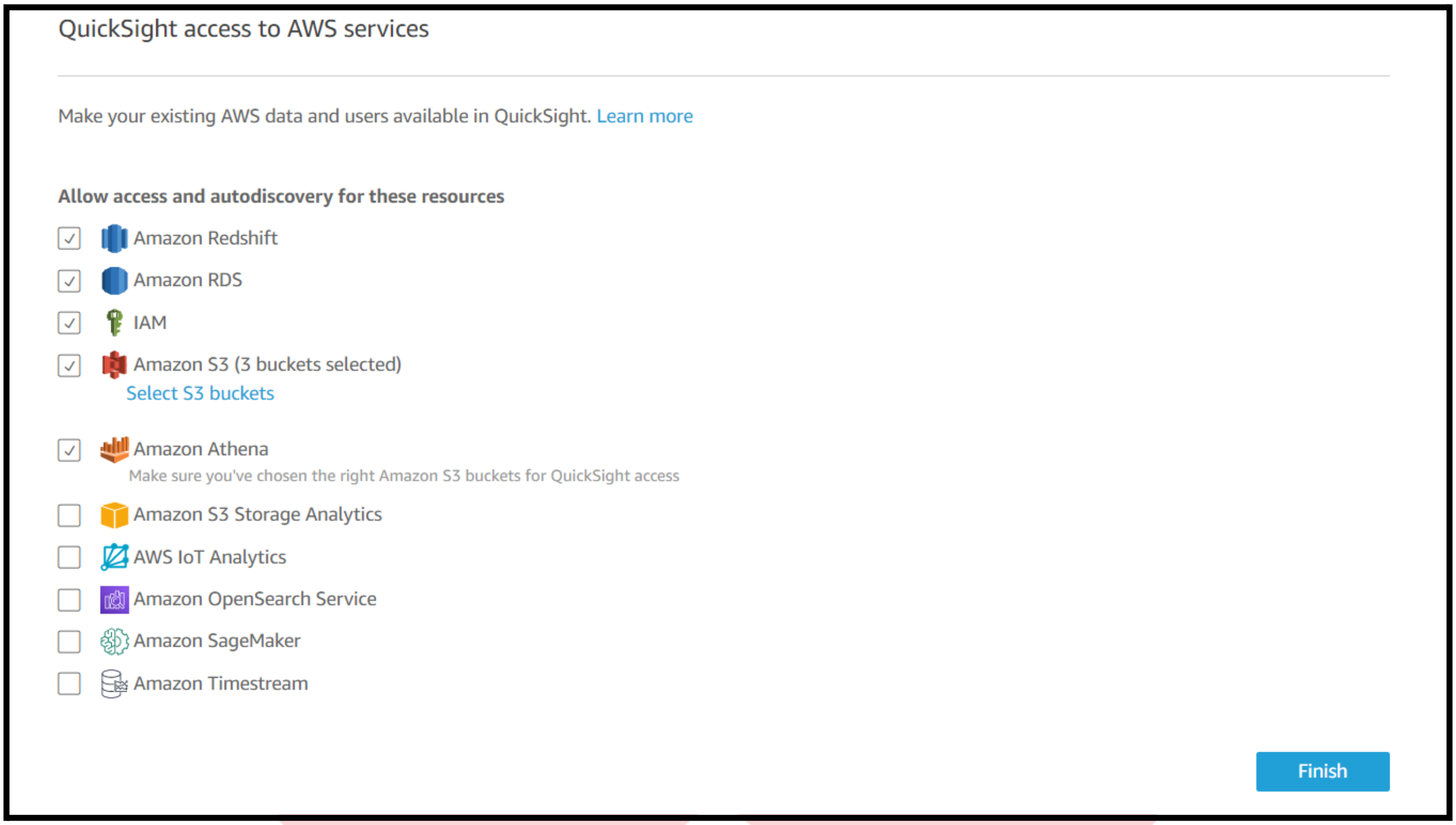
Select the buckets that you want QuickSight to be able to access.

Selected buckets have read only permissions by default. However, you must give write permissions for Athena Workgroup feature.

■ Select all

	S3 Bucket	Write permission for Athena Workgroup
□		□
□		□
☑	data-lake-coal-layer	□
☑	data-lake-diamond-layer	□
☑	data-lake-melting-layer	□
□	data-lake-resources	□
□		□

Check the following services. Make sure your account has enough permissions. This step creates an IAM role for Quicksight.

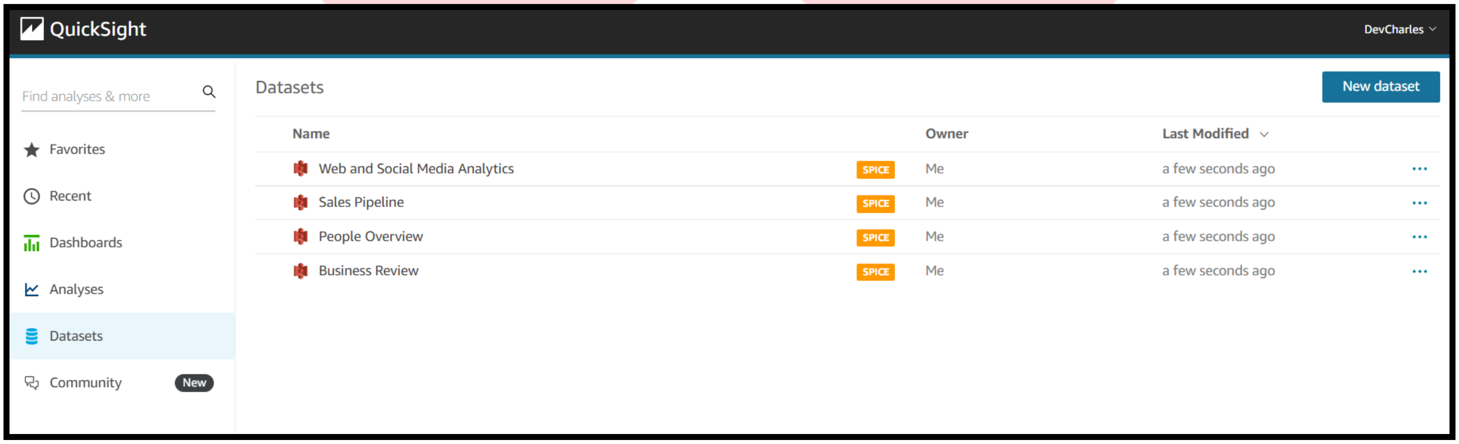


QuickSight access to AWS services

Make your existing AWS data and users available in QuickSight. Learn more

Allow access and autodiscovery for these resources

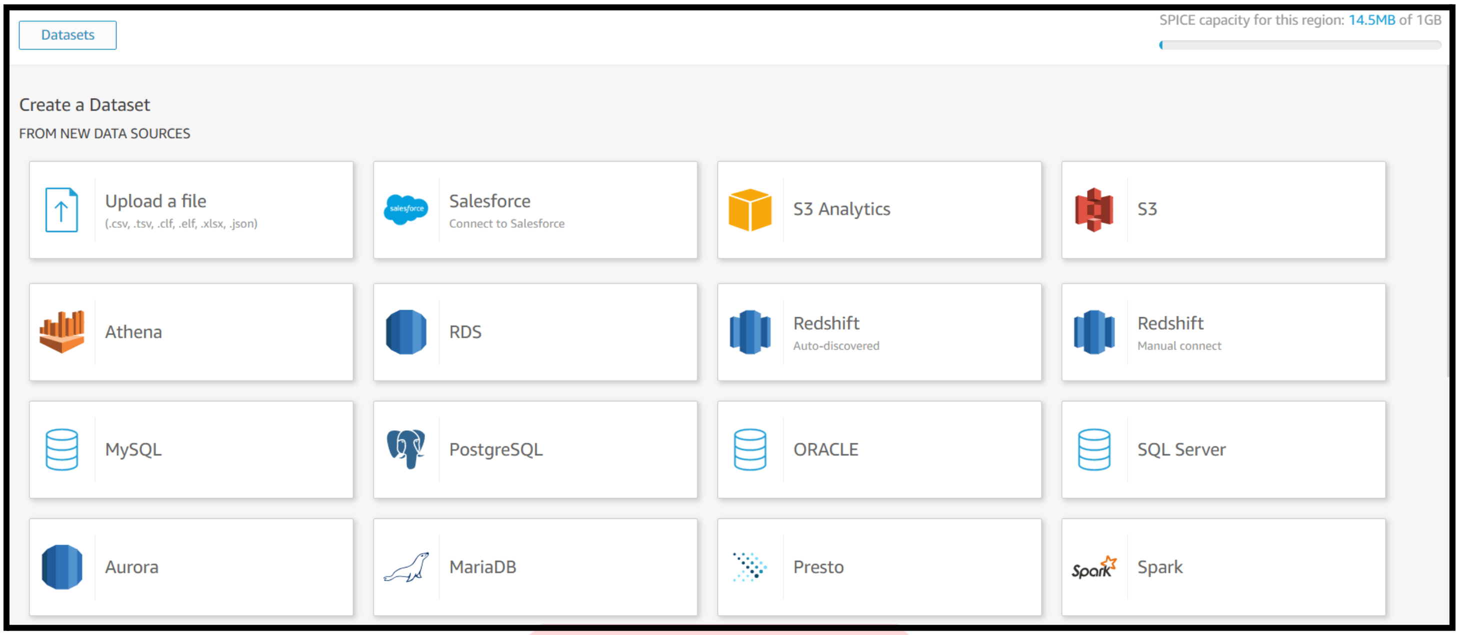
- ☑ Amazon Redshift
- ☑ Amazon RDS
- ☑ IAM
- ☑ Amazon S3 (3 buckets selected)
  Select S3 buckets
- ☑ Amazon Athena
  Make sure you've chosen the right Amazon S3 buckets for QuickSight access
- ☐ Amazon S3 Storage Analytics
- ☐ AWS IoT Analytics
- ☐ Amazon OpenSearch Service
- ☐ Amazon SageMaker
- ☐ Amazon Timestream

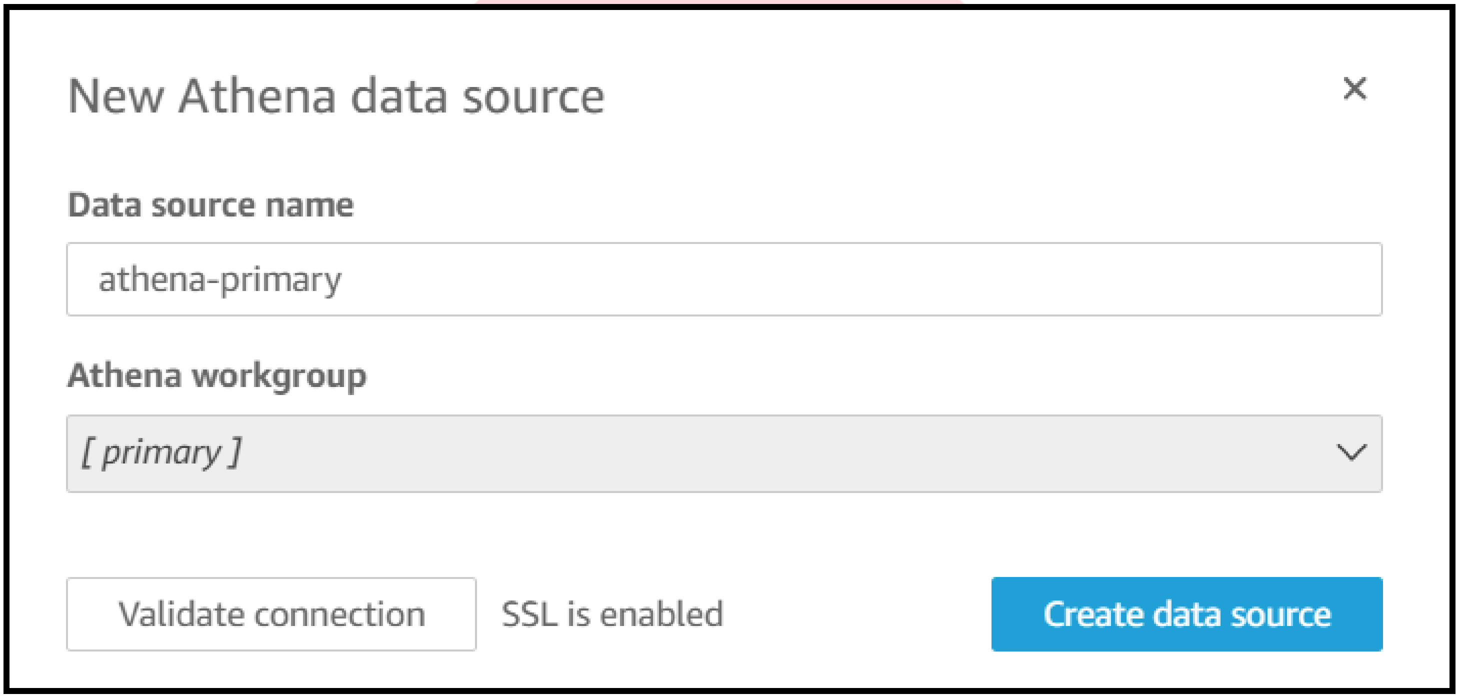


Datasets

Name		Owner	Last Modified
Web and Social Media Analytics	SPICE	Me	a few seconds ago
Sales Pipeline	SPICE	Me	a few seconds ago
People Overview	SPICE	Me	a few seconds ago
Business Review	SPICE	Me	a few seconds ago

Now we will create a data source using Athena. That dataset will get information from our diamond layer.





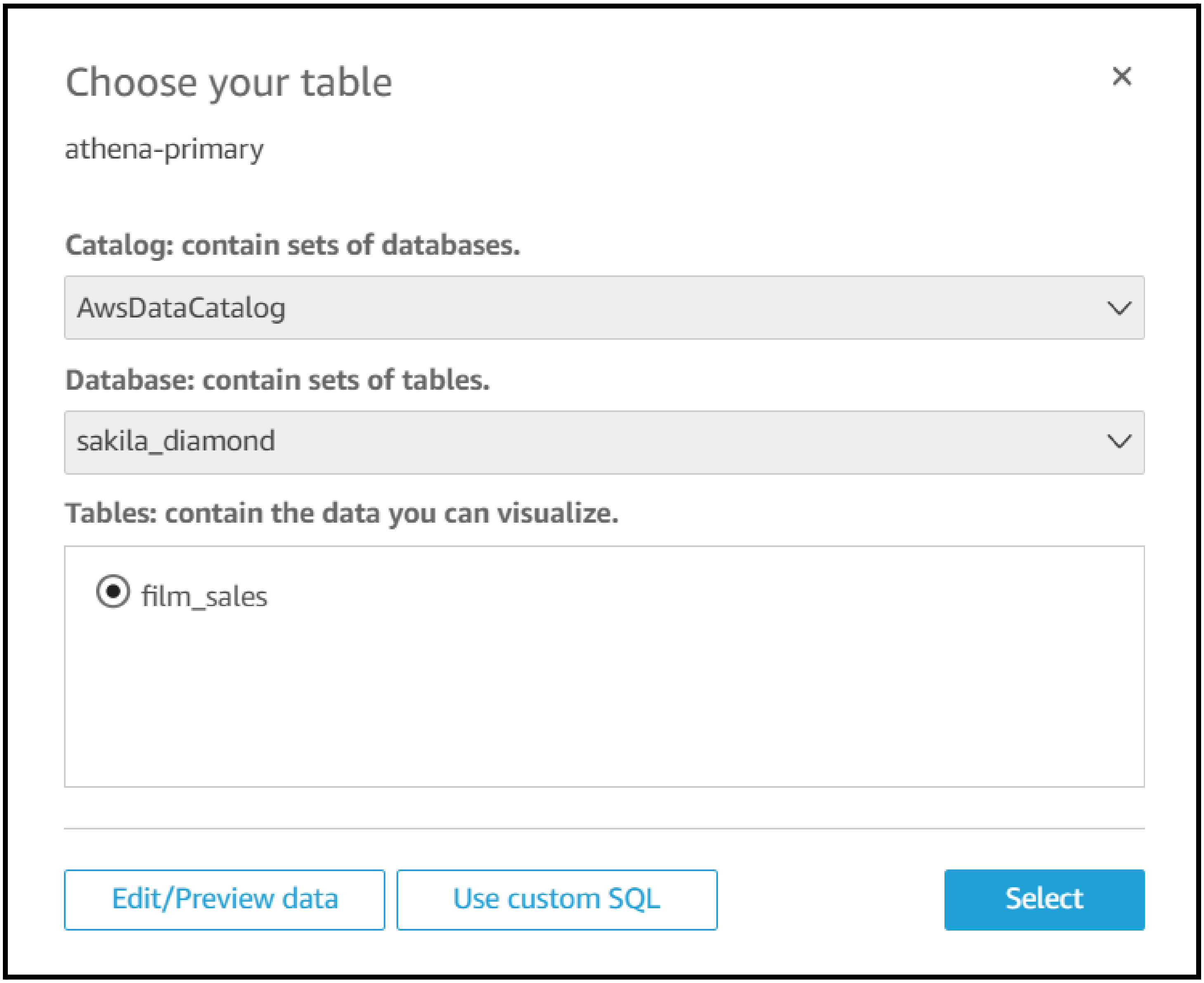
Data source name

athena-primary

Athena workgroup

[primary]

We click on “Edit/Preview data” so we can make sure the data is loaded correctly.



Choose your table

athena-primary

Catalog: contain sets of databases.

AwsDataCatalog

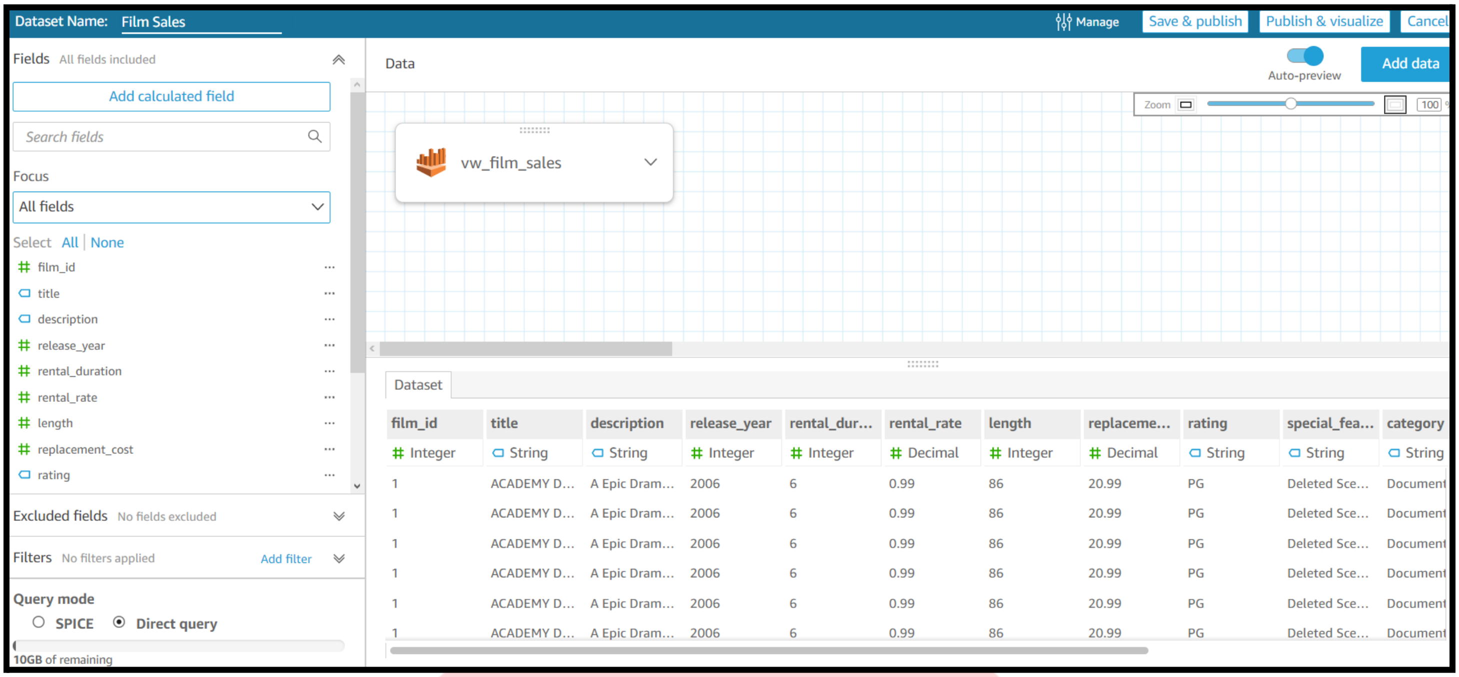
Database: contain sets of tables.

sakila_diamond

Tables: contain the data you can visualize.

● film_sales

Now we set a name for our dataset and click on “Save & visualize”.



Dataset Name: Film Sales

Fields All fields included

Focus: All fields

Select All | None

- film_id
- title
- description
- release_year
- rental_duration
- rental_rate
- length
- replacement_cost
- rating

Excluded fields No fields excluded

Filters No filters applied

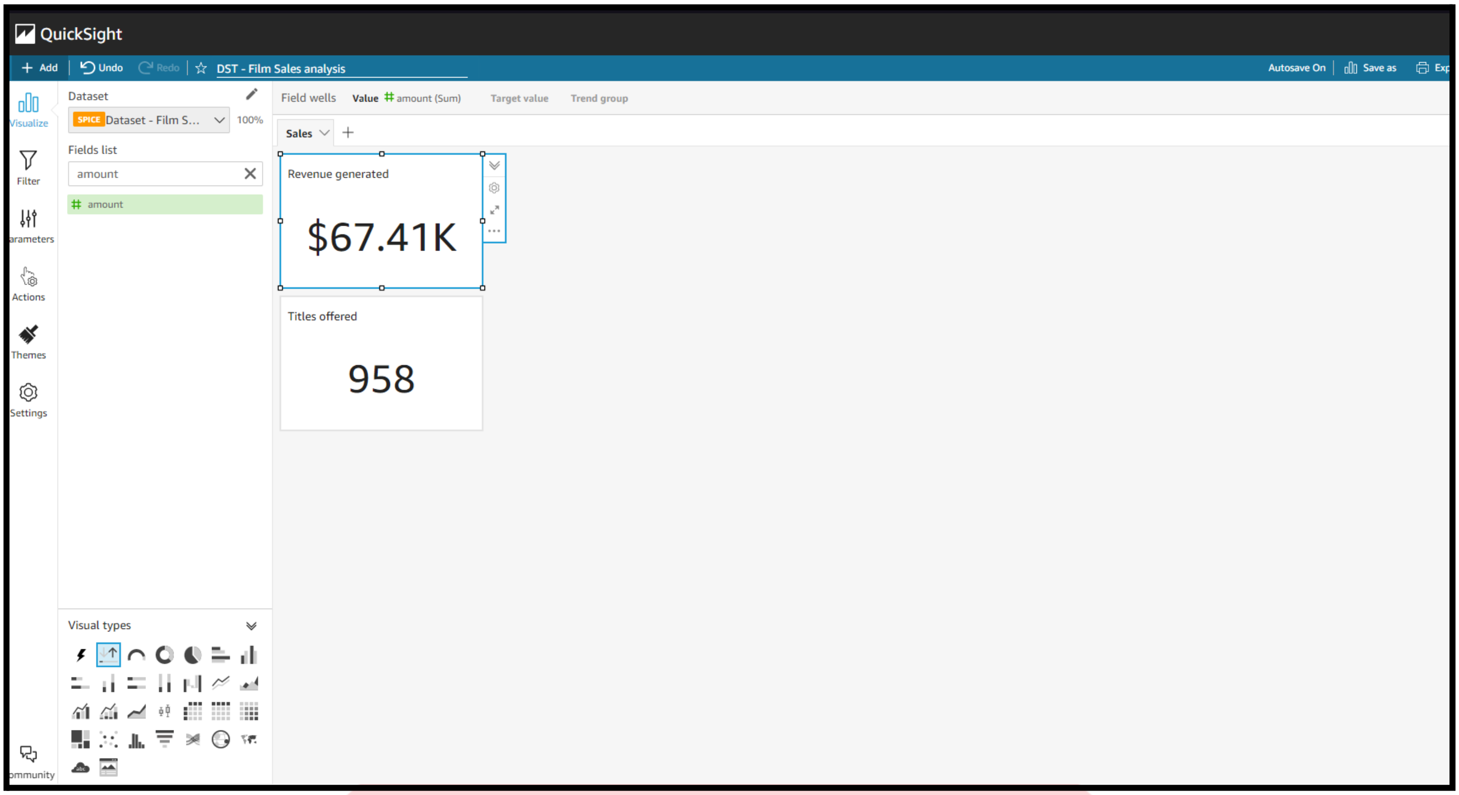
Query mode

○ SPICE ● Direct query

10GB of remaining

film_id	title	description	release_year	rental_dur...	rental_rate	length	replaceme...	rating	special_fea...	category
# Integer	String	String	# Integer	# Integer	# Decimal	# Integer	# Decimal	String	String	String
1	ACADEMY D...	A Epic Dram...	2006	6	0.99	86	20.99	PG	Deleted Sce...	Document
1	ACADEMY D...	A Epic Dram...	2006	6	0.99	86	20.99	PG	Deleted Sce...	Document
1	ACADEMY D...	A Epic Dram...	2006	6	0.99	86	20.99	PG	Deleted Sce...	Document
1	ACADEMY D...	A Epic Dram...	2006	6	0.99	86	20.99	PG	Deleted Sce...	Document
1	ACADEMY D...	A Epic Dram...	2006	6	0.99	86	20.99	PG	Deleted Sce...	Document
1	ACADEMY D...	A Epic Dram...	2006	6	0.99	86	20.99	PG	Deleted Sce...	Document

And with that, we are done. We can create visualizations for our data.


QuickSight
Add
Undo
Redo
DST - Film Sales analysis
Autosave On
Save as
Exp
Visualize
Filter
arameters
Actions
Themes
Settings
ommunity
Dataset
SPICE Dataset - Film S...
100%
Fields list
amount
amount
Visual types
Field wells
Value # amount (Sum)
Target value
Trend group
Sales
Revenue generated
$67.41K
Titles offered
958