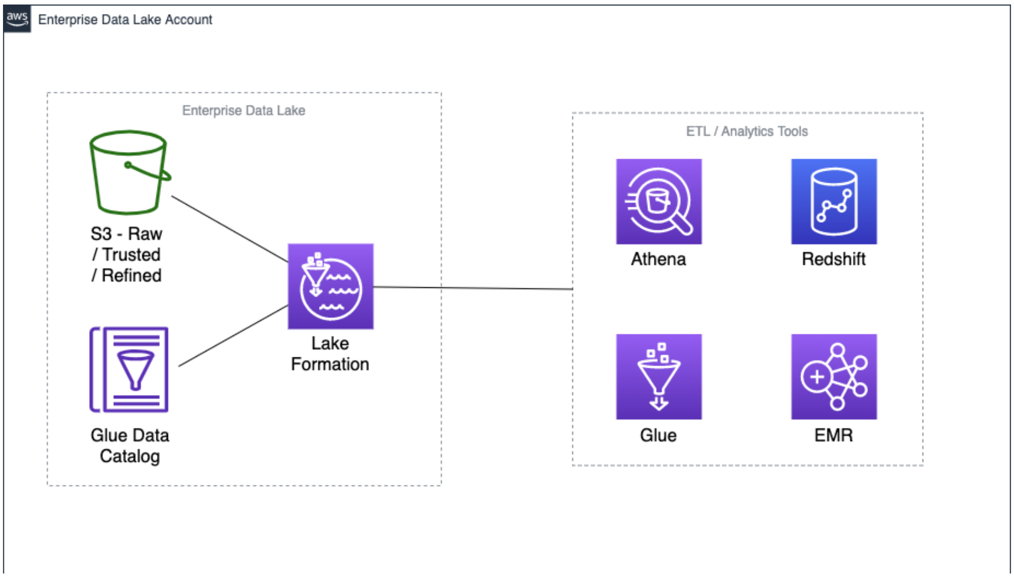
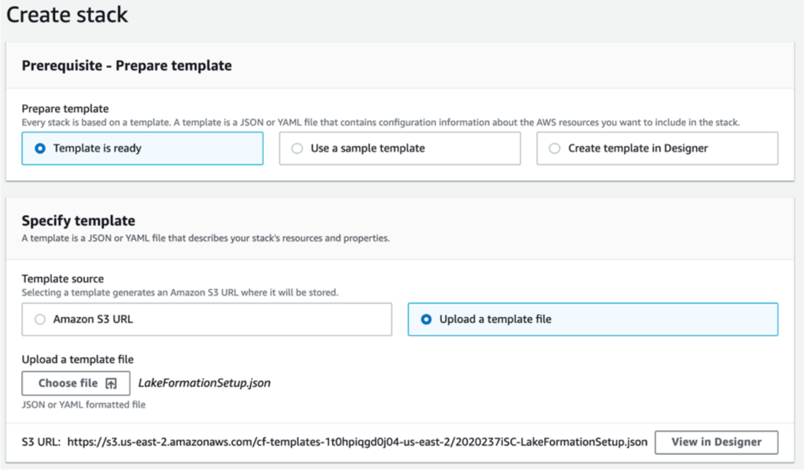
Configuring Lake Formation in a Single Account



1. Go to the data lake account;
2. Create an admin IAM user;
3. Log in as admin;
4. Run CloudFormation (LakeFormationSetup.json);



Create stack

Prerequisite - Prepare template

Prepare template
Every stack is based on a template. A template is a JSON or YAML file that contains configuration information about the AWS resources you want to include in the stack.

● Template is ready ○ Use a sample template ○ Create template in Designer

Specify template
A template is a JSON or YAML file that describes your stack's resources and properties.

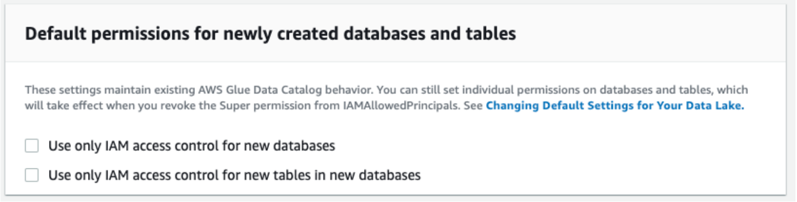
Template source
Selecting a template generates an Amazon S3 URL where it will be stored.

○ Amazon S3 URL ● Upload a template file

Upload a template file
Choose file *LakeFormationSetup.json*
JSON or YAML formatted file

S3 URL: https://s3.us-east-2.amazonaws.com/cf-templates-1t0hpiqgd0j04-us-east-2/2020237iSC-LakeFormationSetup.json

5. Wait until the CloudFormation finishes, then go to Lake Formation, Settings, clear both check boxes and choose Save:

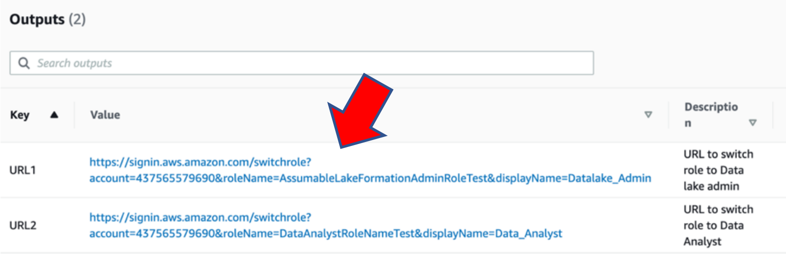


Default permissions for newly created databases and tables

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

6. Go to CloudFormation;
7. Select your stack, see outputs, click on URL1:

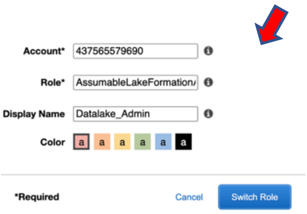


Outputs (2)

Search outputs

Key	Value	Description
URL1	https://signin.aws.amazon.com/switchrole?account=437565579690&roleName=AssumableLakeFormationAdminRoleTest&displayName=Datalake_Admin	URL to switch role to Data lake admin
URL2	https://signin.aws.amazon.com/switchrole?account=437565579690&roleName=DataAnalystRoleNameTest&displayName=Data_Analyst	URL to switch role to Data Analyst

8. Now click “switch role” to assume role Datalake Admin:



Account* 437565579690

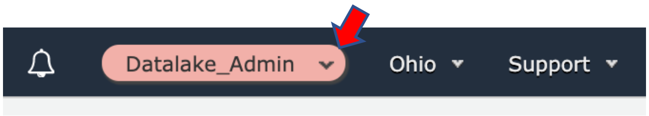
Role* AssumableLakeFormation/

Display Name Datalake_Admin

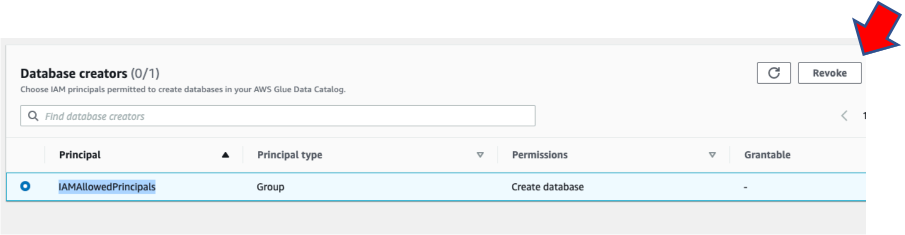
Color a a a a a a

*Required

9. Be sure you are logged as DataLake_Admin (see top console, left side of the selected AWS region):



10. Go to Lake Formation, click “Admins and database creators”, revoke the access to the IAMAllowedPrincipals group:



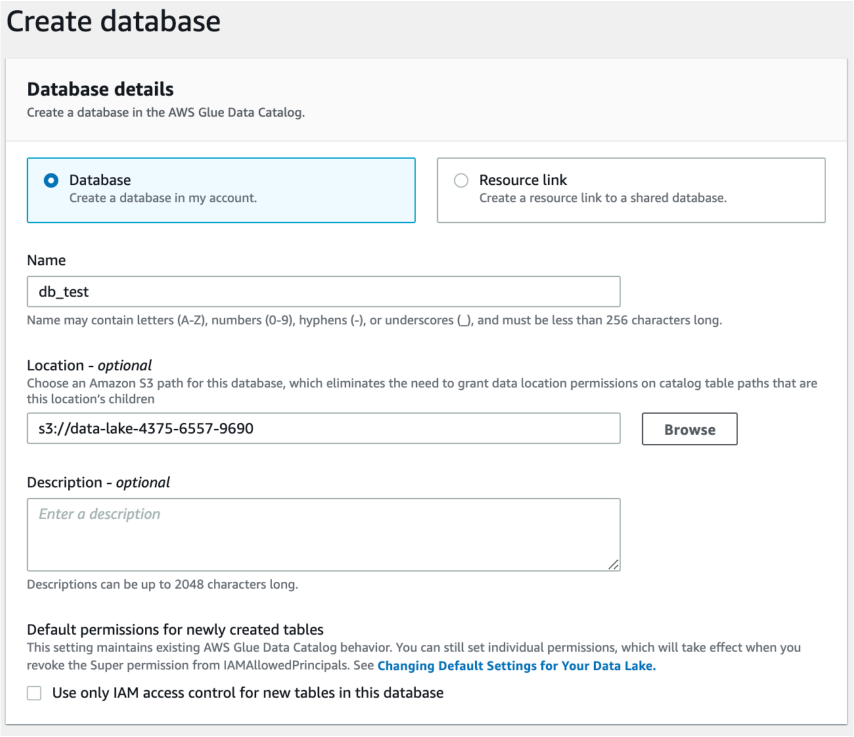
Database creators (0/1)

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

	Principal	Principal type	Permissions	Grantable
●	IAMAllowedPrincipals	Group	Create database	-

11. Create a S3 bucket for your data lake (for testing purposes), and upload some csv files. Also create an additional S3 bucket to store the results of your athena queries.;

12. Go to Lake Formation and create a database called test_db, informing the s3 location:



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

db_test

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

Location - *optional*

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

s3://data-lake-4375-6557-9690

Description - *optional*

Enter a description

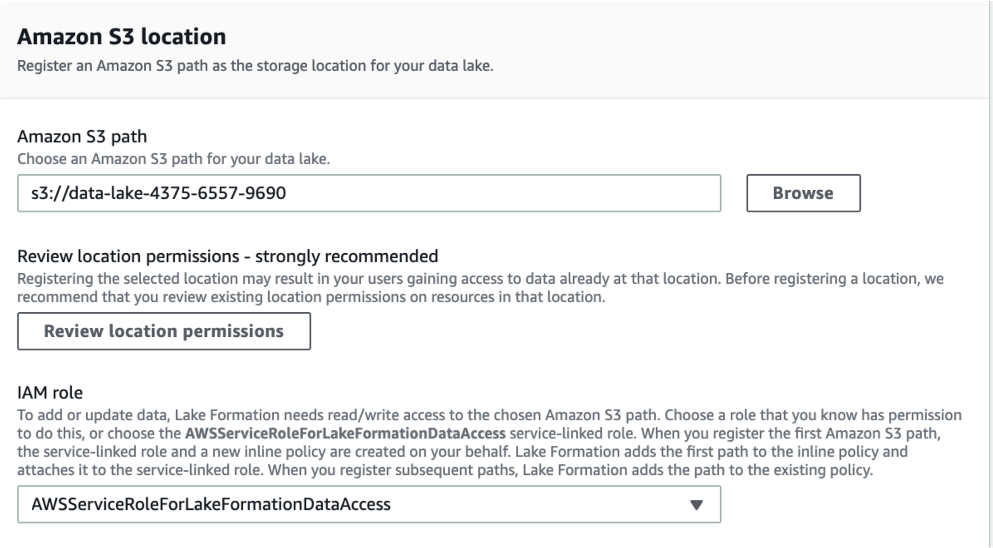
Descriptions can be up to 2048 characters long.

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

13. Go to Data Lake Locations and register your data lake S3 bucket:



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://data-lake-4375-6557-9690

Review location permissions - strongly recommended

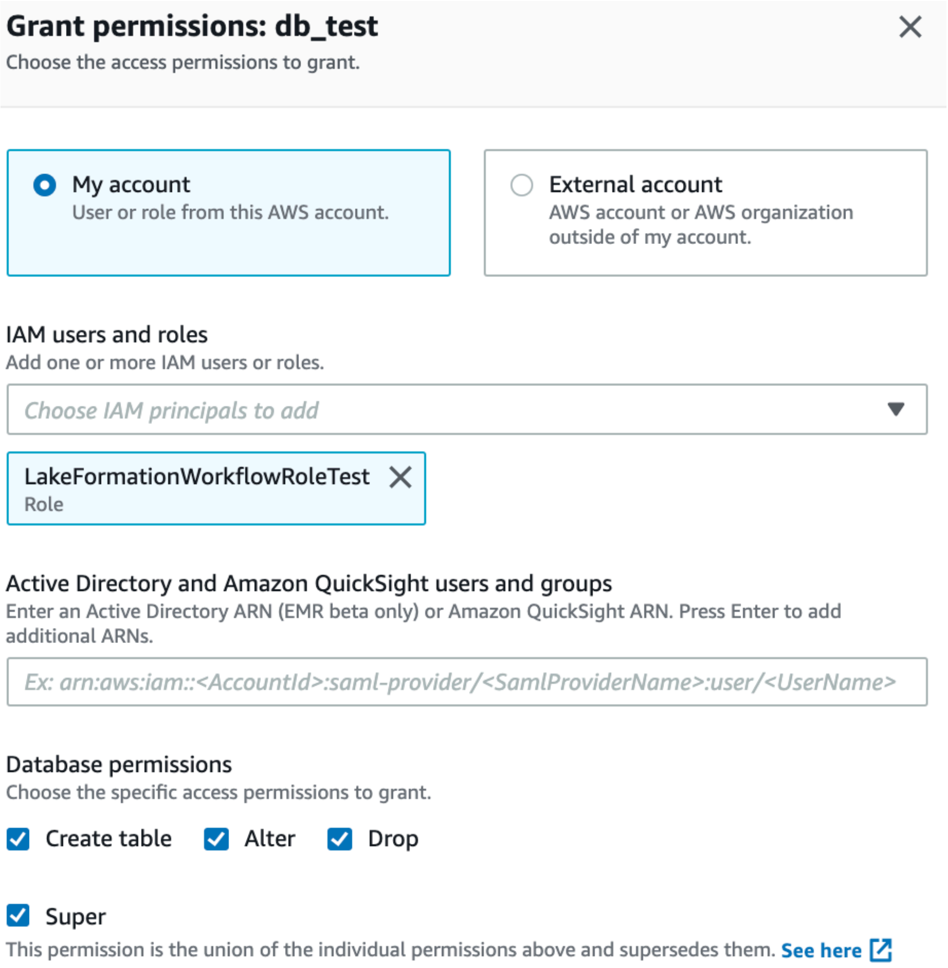
Registering the selected location may result in your users gaining access to data already at that location. Before registering a location, we recommend that you review existing location permissions on resources in that location.

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 ▼

14. Go to Data permissions, and grant access to test_db to the LakeFormationWorkflowRoleTest:



Grant permissions: db_test

Choose the access permissions to grant.

● My account
User or role from this AWS account.

○ External account
AWS account or AWS organization outside of my account.

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

Choose IAM principals to add

LakeFormationWorkflowRoleTest
Role

Active Directory and Amazon QuickSight users and groups
Enter an Active Directory ARN (EMR beta only) or Amazon QuickSight ARN. Press Enter to add additional ARNs.

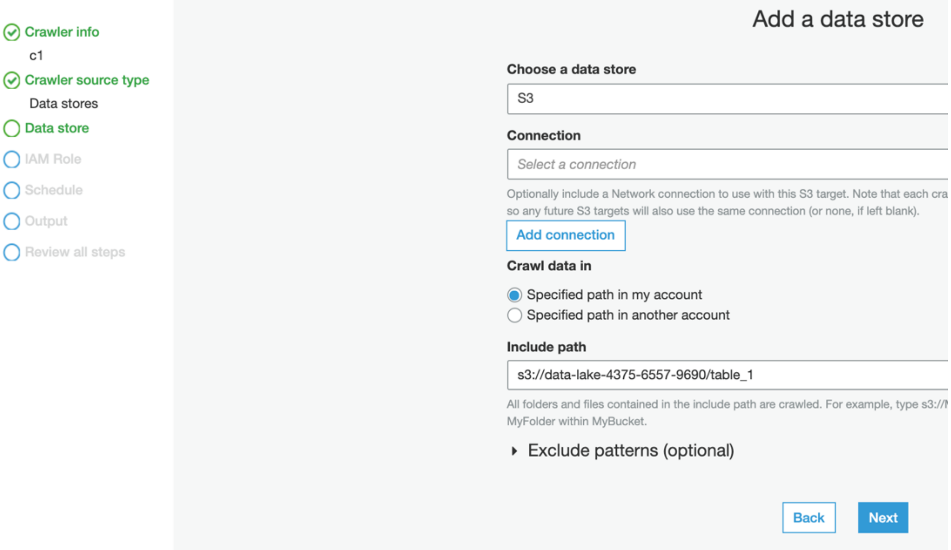
Ex: arn:aws:iam::<AccountId>:saml-provider/<SamlProviderName>:user/<UserName>

Database permissions
Choose the specific access permissions to grant.

☑ Create table ☑ Alter ☑ Drop

☑ Super
This permission is the union of the individual permissions above and supersedes them. See here

15. Go to Glue and create a crawler to create some tables in your data lake:



Add a data store

Choose a data store
S3

Connection
Select a connection

Optionally include a Network connection to use with this S3 target. Note that each cra
so any future S3 targets will also use the same connection (or none, if left blank).

Add connection

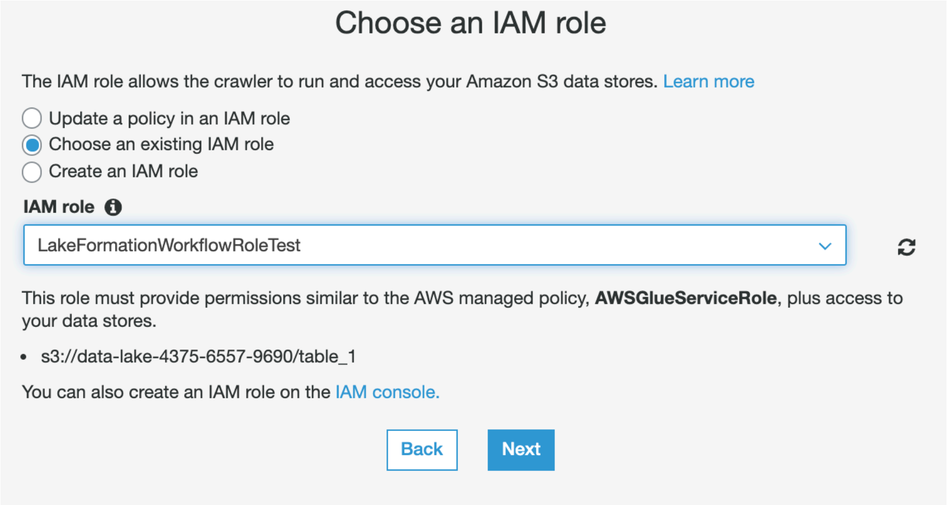
Crawl data in
● Specified path in my account
○ Specified path in another account

Include path
s3://data-lake-4375-6557-9690/table_1

All folders and files contained in the include path are crawled. For example, type s3://M
MyFolder within MyBucket.

▸ Exclude patterns (optional)

16. Choose the role LakeFormationWorkflowRoleTest:



Choose an IAM role

The IAM role allows the crawler to run and access your Amazon S3 data stores. Learn more

○ Update a policy in an IAM role
● Choose an existing IAM role
○ Create an IAM role

IAM role

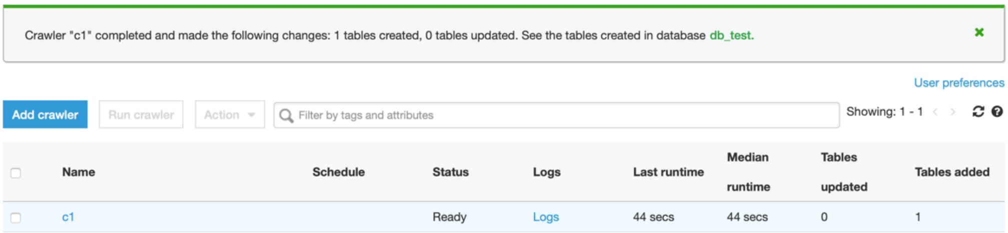
LakeFormationWorkflowRoleTest

This role must provide permissions similar to the AWS managed policy, **AWSGlueServiceRole**, plus access to your data stores.

- s3://data-lake-4375-6557-9690/table_1

You can also create an IAM role on the IAM console.

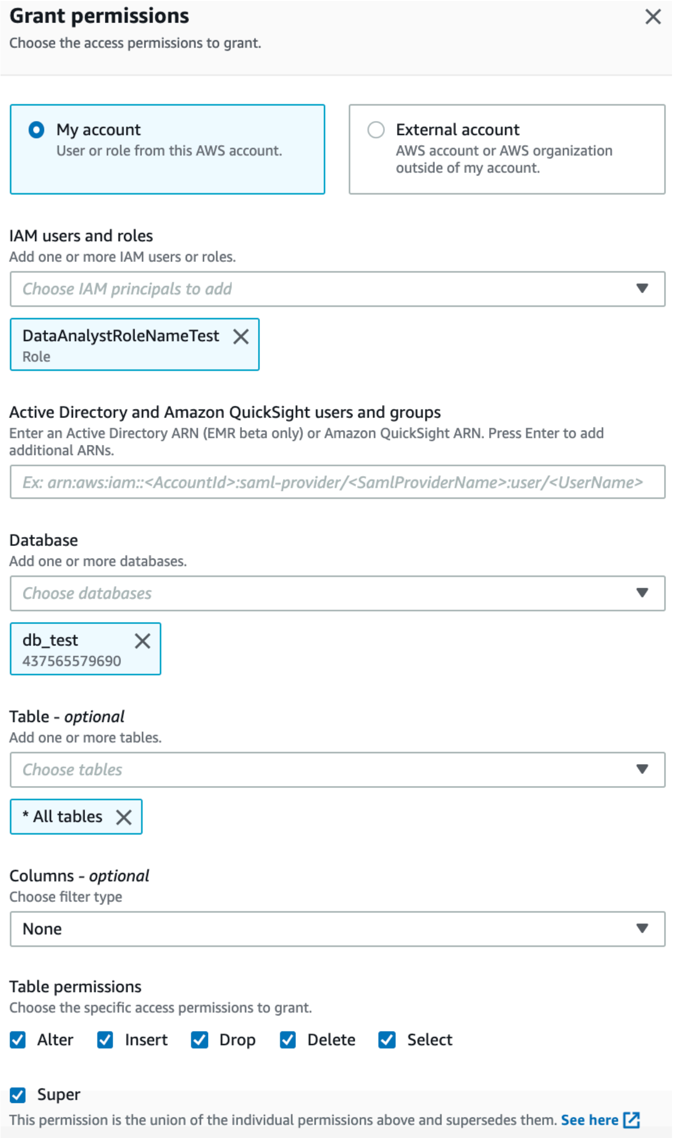
17. Run your crawler. It should create at least one table in order to allow us to do some tests:



Crawler "c1" completed and made the following changes: 1 tables created, 0 tables updated. See the tables created in database db_test.

□	Name	Schedule	Status	Logs	Last runtime	Median runtime	Tables updated	Tables added
□	c1		Ready	Logs	44 secs	44 secs	0	1

18. Now that you have created your first table under Lake Formation security rules, it is time to grant access to the Data Analyst Role to all tables in the db_test database:



Grant permissions

Choose the access permissions to grant.

● My account
User or role from this AWS account.

○ External account
AWS account or AWS organization outside of my account.

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

DataAnalystRoleNameTest
Role

Active Directory and Amazon QuickSight users and groups
Enter an Active Directory ARN (EMR beta only) or Amazon QuickSight ARN. Press Enter to add additional ARNs.
Ex: arn:aws:iam::<AccountId>:saml-provider/<SamlProviderName>:user/<UserName>

Database
Add one or more databases.
Choose databases

db_test
437565579690

Table - *optional*
Add one or more tables.
Choose tables

* All tables

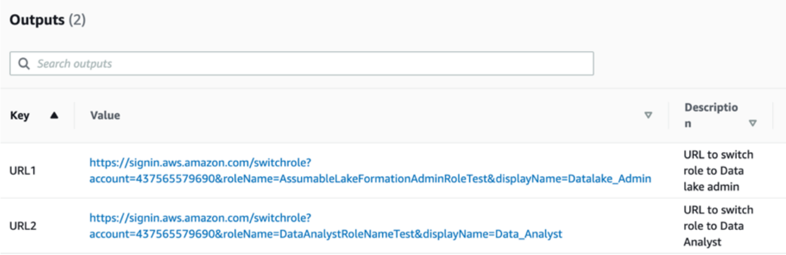
Columns - *optional*
Choose filter type
None

Table permissions
Choose the specific access permissions to grant.

☑ Alter ☑ Insert ☑ Drop ☑ Delete ☑ Select

☑ Super
This permission is the union of the individual permissions above and supersedes them. See here

19. Now go to CloudFormation, click in the URL2 link to assume the DataAnalyst role:

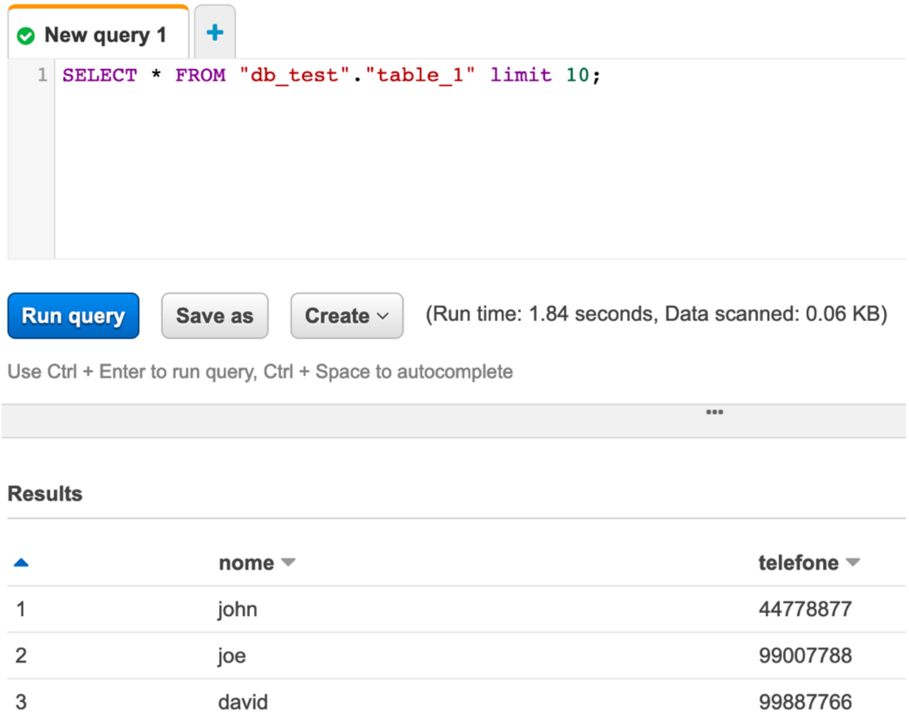


Outputs (2)

Search outputs

Key	Value	Description
URL1	https://signin.aws.amazon.com/switchrole?account=437565579690&roleName=AssumableLakeFormationAdminRoleTest&displayName=Datalake_Admin	URL to switch role to Data lake admin
URL2	https://signin.aws.amazon.com/switchrole?account=437565579690&roleName=DataAnalystRoleNameTest&displayName=Data_Analyst	URL to switch role to Data Analyst

20. Go to Athena, setup a query result location in S3, and query your test table. Use your admin account to give access to a S3 bucket to the Data Analyst role, so you can write the results to a S3 bucket:



```
SELECT * FROM "db_test"."table_1" limit 10;
```

(Run time: 1.84 seconds, Data scanned: 0.06 KB)

Use Ctrl + Enter to run query, Ctrl + Space to autocomplete

Results

	nome	telefone
1	john	44778877
2	joe	99007788
3	david	99887766