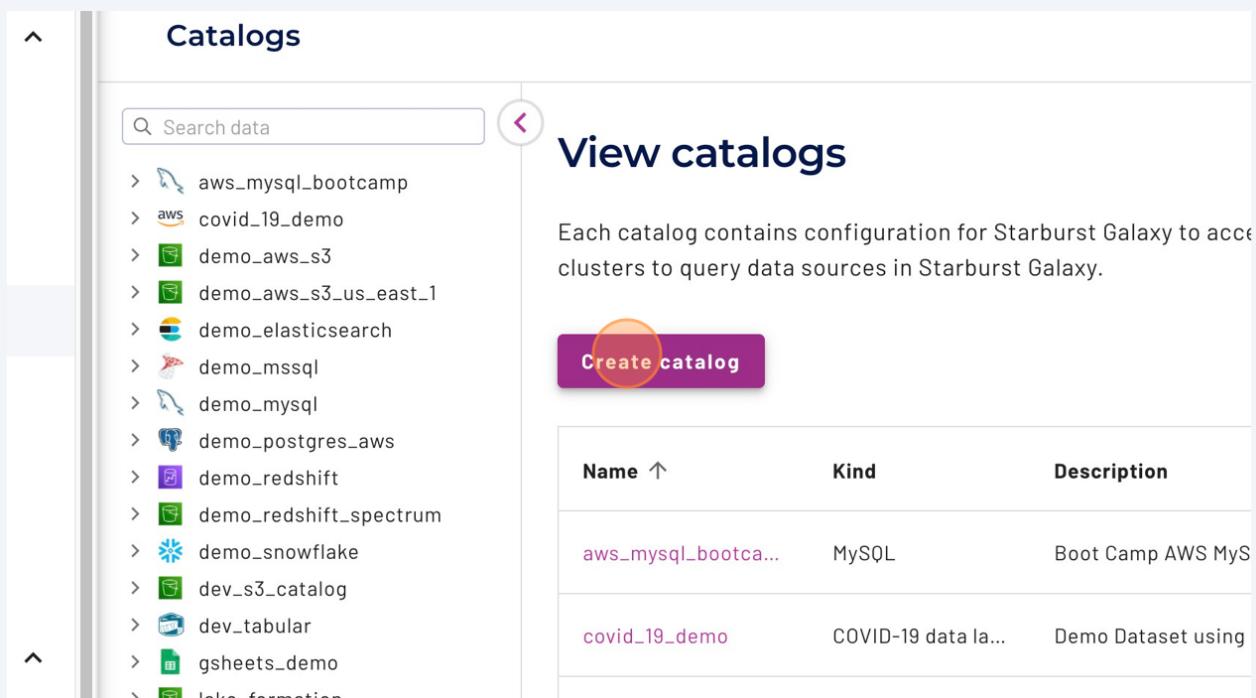


Creating a Catalog in Starburst Galaxy Configuration Hub

Scribe 

- 1 Navigate to ycat.galaxy.starburst.io/catalog

- 2 Click "Create catalog"



Catalogs

Search data

- >  aws_mysql_bootcamp
- >  covid_19_demo
- >  demo_aws_s3
- >  demo_aws_s3_us_east_1
- >  demo_elasticsearch
- >  demo_mssql
- >  demo_mysql
- >  demo_postgres_aws
- >  demo_redshift
- >  demo_redshift_spectrum
- >  demo_snowflake
- >  dev_s3_catalog
- >  dev_tabular
- >  gsheets_demo
- >  lakeFormation

View catalogs

Each catalog contains configuration for Starburst Galaxy to access clusters to query data sources in Starburst Galaxy.

Create catalog

Name ↑	Kind	Description
aws_mysql_bootca...	MySQL	Boot Camp AWS MyS...
covid_19_demo	COVID-19 data la...	Demo Dataset using...

- 3 Click this image.

4 Set permissions (optional)

4 Add to cluster (optional)

Cloud object storage

These cloud object storage catalogs support **Iceberg**, **Hive**, **DeltaLake** added to an accelerated cluster for faster querying. Learn about [Warp](#)



Amazon S3



Azure Data Lake Storage

Additional data sources



Amazon Redshift



Azure Synapse

- 4 Click here.

- 4 Configure the connection
- 3 Set permissions (optional)
- 4 Add to cluster (optional)

Configure your catalog to query objects in Amazon S3. Learn more about [connecting to S3](#).

Name and description

Provide a unique name to identify the catalog in your SQL queries in the query editor and other client tools. The namespace for a table is typically <catalog_name>.<schema_name>.<table_name>

Catalog name * ... ?

Must start with a letter and only use lowercase letters (a-z), numbers (0-9), and underscores (_)

Description ?

Authentication to S3

Choose the authentication mechanism ? to connect to S3.

Authentication with *

Cross account IAM role AWS access key

5 Click the "Catalog name *" field.

2 Configure the connection

Configure your catalog to query objects in Amazon S3. Learn more about [connecting to S3](#).

3 Set permissions (optional)

4 Add to cluster (optional)

Name and description

Provide a unique name to identify the catalog in your SQL queries in the query editor and other client tools. The namespace for a table is typically <catalog_name>.<schema_name>.<table_name>

Catalog name *

Must start with a letter and only use lowercase letters (a-z), numbers (0-9), and underscores (_)

Description

Authentication to S3

Choose the authentication mechanism to connect to S3.

Authentication with *

6 Click the "AWS access key" field.

Must start with a letter and only use lowercase letters (a-z), numbers (0-9), and underscores (_)

Description

example object

Authentication to S3

Choose the authentication mechanism to connect to S3.

Authentication with *

Cross account IAM role AWS access key

Cross account IAM role *

+ Configure a cross account IAM role

Metastore configuration

Configure access to the metastore to provide metadata and mapping

- 7 Click the "AWS access key for S3 *" field.

Description
example object ?

Authentication to S3

Choose the authentication mechanism  to connect to S3.

Authentication with *

Cross account IAM role AWS access key

AWS access key for S3 * ... ?

AWS secret key for S3 * ... eye icon ?

Metastore configuration

Configure access to the metastore to provide metadata and mapping information about the objects stored in Amazon S3.

Metastore type *

- 8 Click the "AWS secret key for S3 *" field.

Authentication to S3

Choose the authentication mechanism  to connect to S3.

Authentication with *

Cross account IAM role AWS access key

AWS access key for S3 * AKIAWAI3RINJH5LWXFW6 ... ?

AWS secret key for S3 * ... eye icon ?

Metastore configuration

Configure access to the metastore to provide metadata and mapping information about the objects stored in Amazon S3.

Metastore type *

AWS Glue Hive Metastore Starburst Galaxy

- 9 Click the "Default S3 bucket name *" field.

Metastore configuration

Configure access to the metastore to provide metadata and mapping information about the objects stored in Amazon S3.

Metastore type *

AWS Glue Hive Metastore Starburst Galaxy

Default S3 bucket name *



Default directory name *



Allow creating external tables



Allow writing to external tables



- 10 Click the "Default directory name *" field.

Configure access to the metastore to provide metadata and mapping information about the objects stored in Amazon S3.

Metastore type *

AWS Glue Hive Metastore Starburst Galaxy

Default S3 bucket name *

yusuf-cattaneo-bootcamp-nov2022



Default directory name *



Allow creating external tables



Allow writing to external tables



Default table format

Select the default table format used for creating new tables. The catalog will be able to read from any type. [Check out our docs](#) to learn more.

- 11 Click the "Allow creating external tables" field.

Metastore type *

AWS Glue Hive Metastore Starburst Galaxy

Default S3 bucket name * ... ?

Default directory name * ?

Allow creating external tables ?

Allow writing to external tables ?

Default table format

Select the default table format used for creating new tables. The catalog will be able to read from any type. [Check out our docs](#) ? to learn more.

Default table format *

- 12 Click the "Allow writing to external tables" field.

AWS Glue Hive Metastore Starburst Galaxy

Default S3 bucket name * ... ?

Default directory name * ?

Allow creating external tables ?

Allow writing to external tables ?

Default table format

Select the default table format used for creating new tables. The catalog will be able to read from any type. [Check out our docs](#) ? to learn more.

Default table format *

Iceberg Hive Delta Lake

13 Click "Test connection"

Default table format

Select the default table format used for creating new tables. The catalog will be able to read from any type. [Check out our docs](#) to learn more.

Default table format *

Iceberg Hive Delta Lake

Test connection

Validate that the network configuration allows Starburst Galaxy to connect to the data source.

Test connection

[Back](#) [Connect catalog](#)

14 Click "Connect catalog"

Iceberg Hive Delta Lake

Test connection

Validate that the network configuration allows Starburst Galaxy to connect to the data source.

Detected regions:

- aws US East (Ohio)

Hooray! You can now add this catalog to a cluster.

Test connection

[Back](#) **Connect catalog**

- 15 Click the "Read-only catalog" field.

- Select a data source
- Configure the connection
- Set permissions (optional)
- Add to cluster (optional)

Set permissions

Now that your **example_object_storage** catalog has been created, assign users access with roles. [Learn how to create roles here.](#)

Catalog-level permissions



Read-only catalog

Prohibits all users, **including the catalog owner**, from modifying data or metadata in this catalog.

Role-level permissions

The following roles will be able to read and write data and metadata in this catalog, including creating and deleting schemas and tables. The specific privileges included are detailed in [the documentation](#).

- 16 Click the "Read-only catalog" field.

- Select a data source
- Configure the connection
- Set permissions (optional)
- Add to cluster (optional)

Set permissions

Now that your **example_object_storage** catalog has been created, assign users access with roles. [Learn how to create roles here.](#)

Catalog-level permissions



Read-only catalog

Prohibits all users, **including the catalog owner**, from modifying data or metadata in this catalog.

Role-level permissions

The following roles will be able to read data and metadata from all schemas and tables within this catalog, as described in [the documentation](#).

Roles with read access

17 Click "Save access controls"

metadata in this catalog, including creating and deleting schemas and tables. The specific privileges included are detailed in [the documentation](#).

Roles with read and write access

accountadmin

The following roles will be able to read data and metadata from all schemas and tables within this catalog, as described in [the documentation](#).

Roles with read access

accountadmin

[Skip](#)

[Save access controls](#)

18 Click the "Select clusters" field.

[elect a data source](#)

[configure the connection](#)

[set permissions \(optional\)](#)

[add to cluster \(optional\)](#)

Add to cluster

Attach your **example_object_storage** catalog to a cluster in order to query your data. You may add it to an existing cluster in the same region, or create a new cluster.

Add to cluster

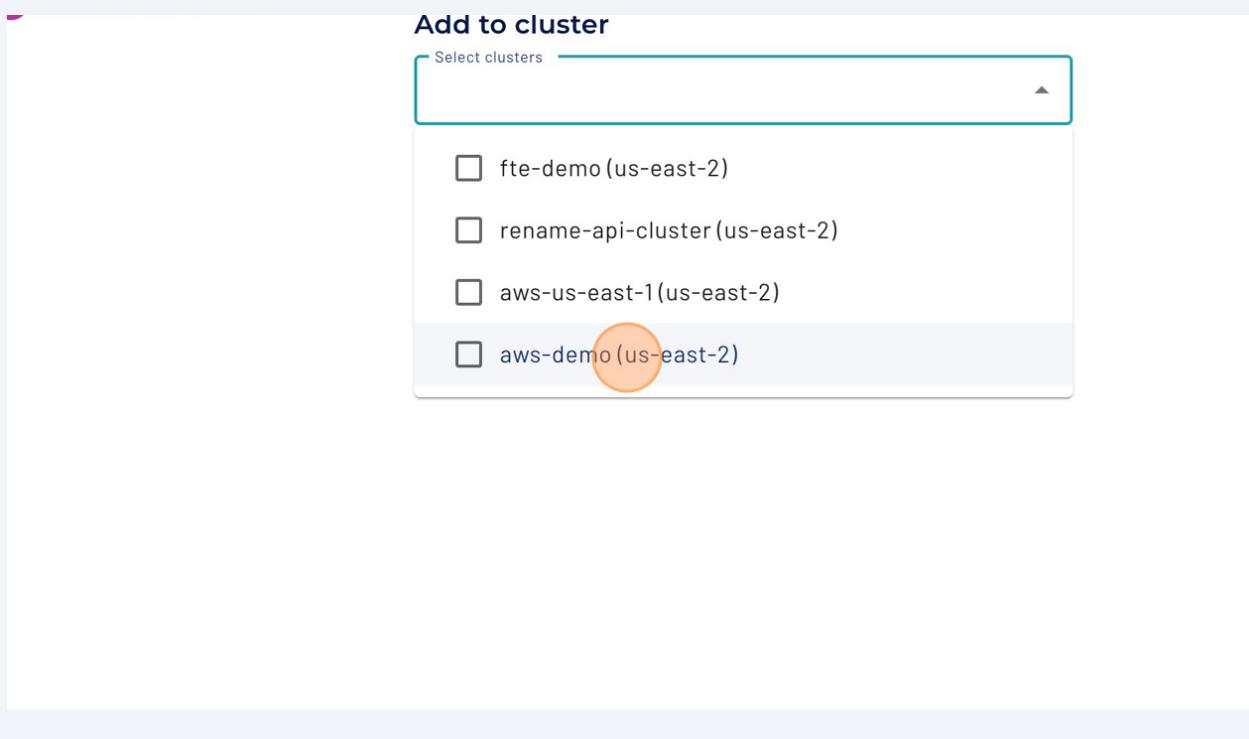
Select clusters

[+ Create a new cluster](#)

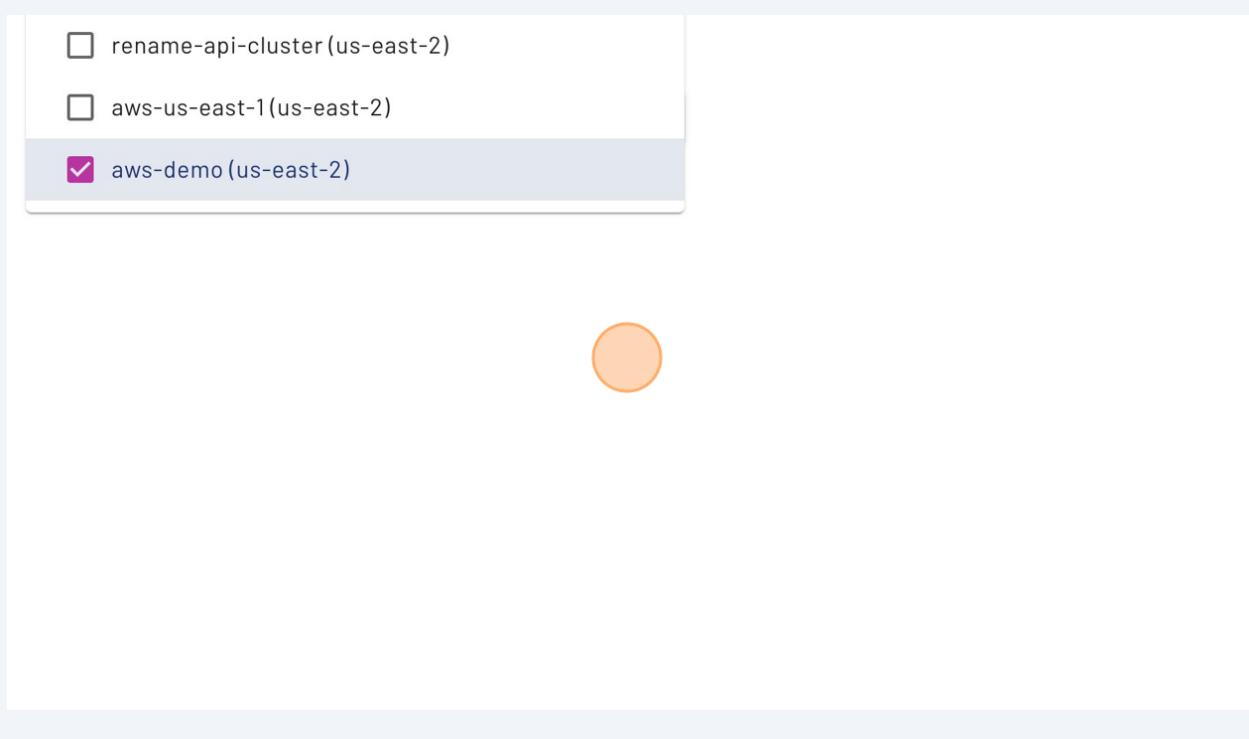
[Skip](#)

[Add to cluster](#)

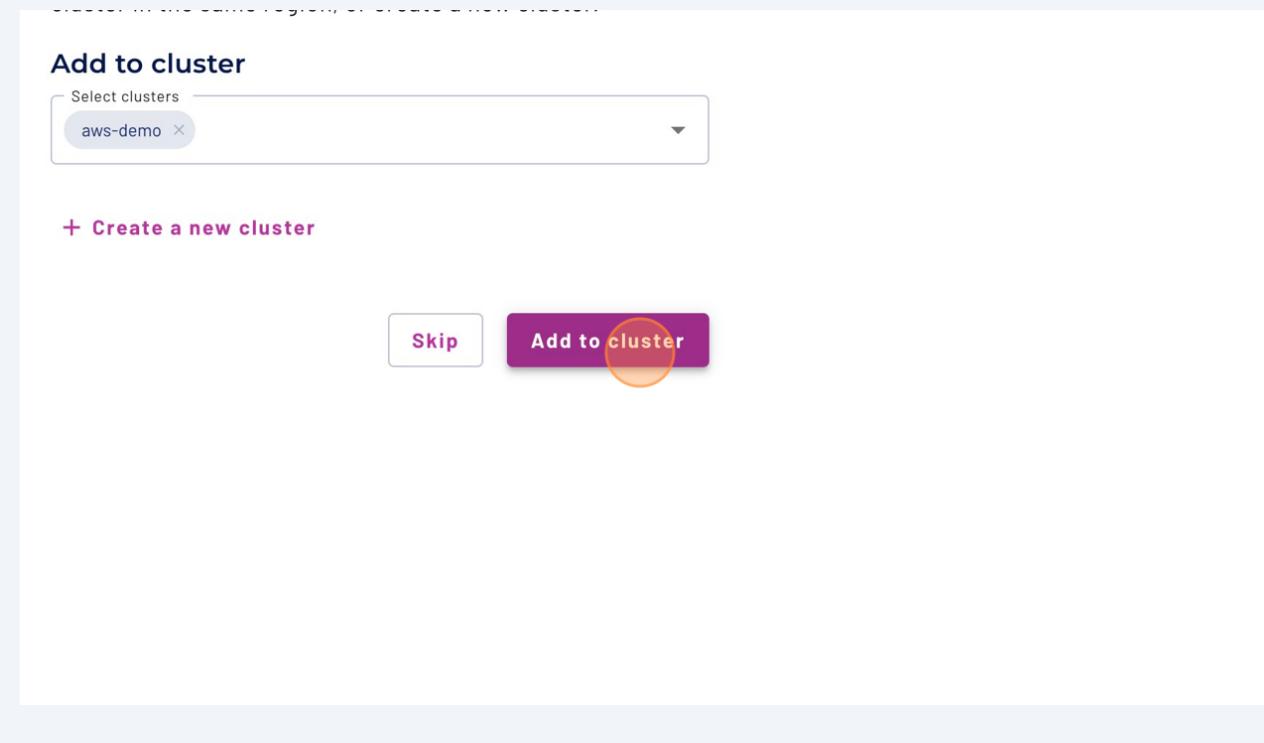
19 Click "aws-demo (us-east-2)"



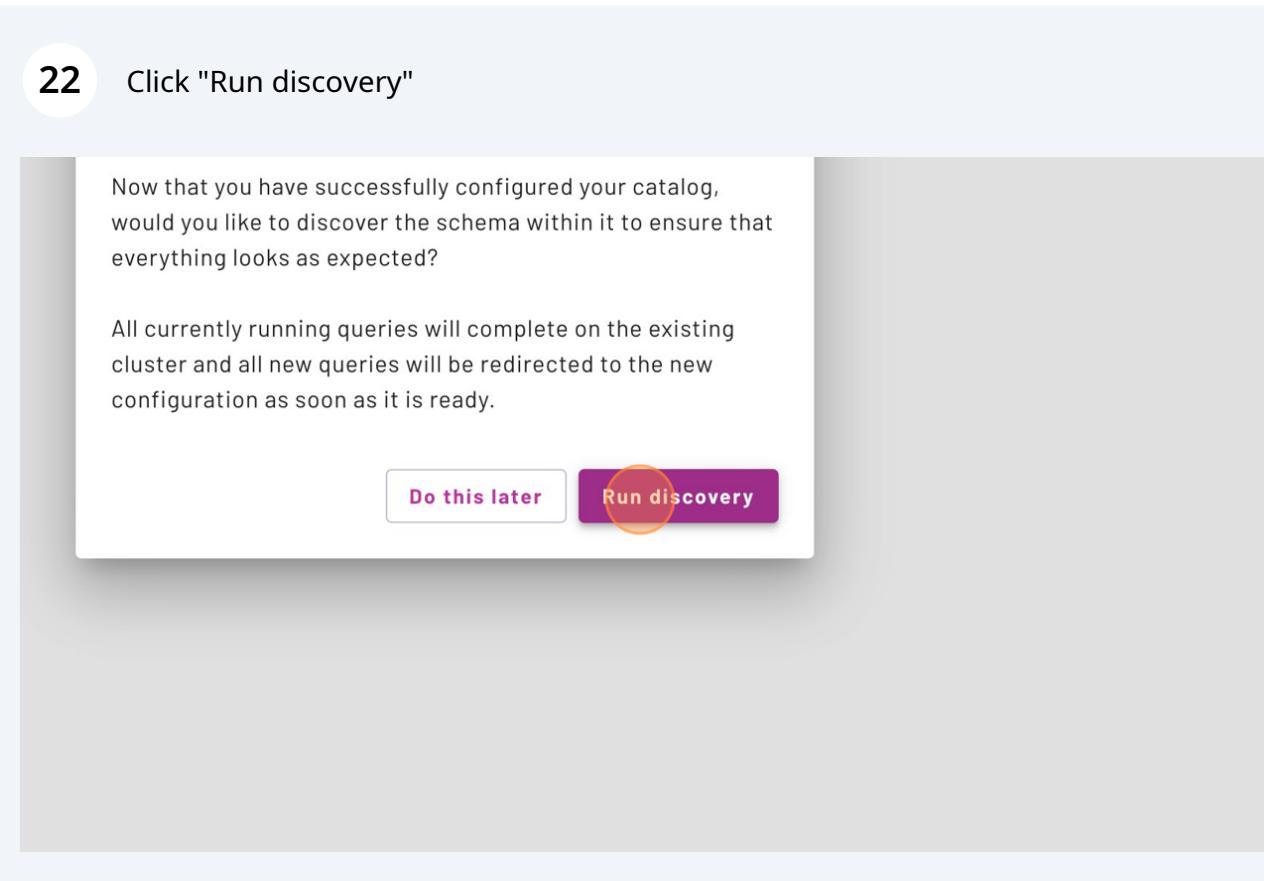
20 Click here.



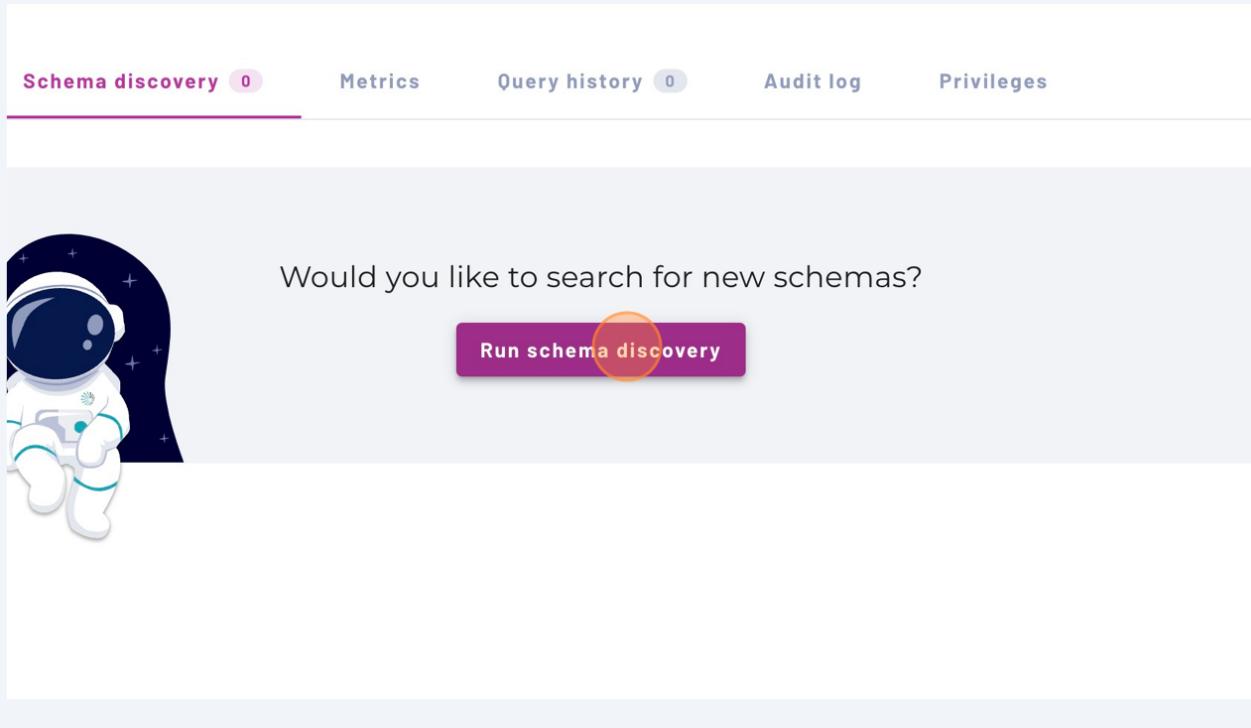
21 Click "Add to cluster"



22 Click "Run discovery"



23 Click "Run schema discovery"



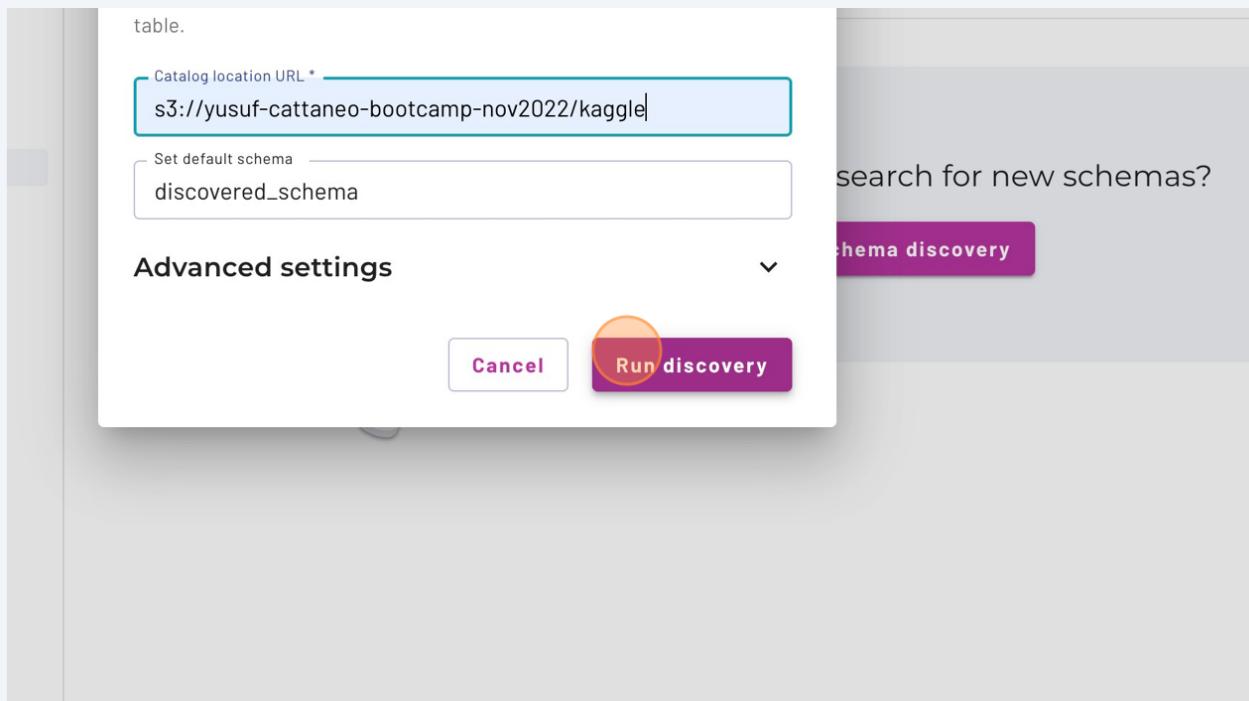
24 Click the "Catalog location URL *" field.

The screenshot shows a 'Run discovery' dialog box. The title is 'Run discovery'. Inside the dialog, there is a text input field labeled 'Catalog location URL *' which is highlighted with a yellow circle. Below the input field is a dropdown menu set to 'discovered_schema'. At the bottom of the dialog are two buttons: 'Cancel' and 'Run discovery' (which is highlighted with a yellow circle).

On the left side of the dialog, there is a sidebar listing various catalog entries:

- > demo_aws_s3
- > demo_aws_s3_us_east_1
- > demo_elasticsearch
- > demo_mssql
- > demo_mysql
- > demo_postgres_aws
- > demo_redshift
- > demo_redshift_spectrum
- > demo_snowflake
- > dev_s3_catalog
- > dev_tabular
- > example_object_storage
- > gsheets_demo
- > lakeFormation
- > personal_postgres
- > s3_bootcamp
- > s3_bootcamp_glue
- > s3_bootcamp_iamrole
- > s3_hive_transactional

25 Click "Run discovery"



26 Click "s3://yusuf-cattaneo-bootcamp-nov2022/kaggle"

A screenshot of the Schema Discovery interface. On the left, there's a sidebar with various database and storage options like "postgres_aws", "redshift", "redshift_spectrum", "snowflake", "s3_catalog", "tabular", and "multiple_object_storage". The "multiple_object_storage" section is expanded, showing "ets_demo", ".formation", "onal_postgres", "bootcamp", "bootcamp_glue", "bootcamp_iamrole", "hive_transactional", "ole", "ole_data", ".hive", and ".no_datalake". The main area has tabs for "Schemas", "Schema discovery" (which is active and has a count of 1), "Metrics", and "Query history". Under the "Schema discovery" tab, there's a "Run discovery" button. Below it is a table with three columns: "Source", "Timestamp", and "Status". There is one entry: "s3://yusuf-cattaneo-bootcamp-nov2022/kaggle" (with a yellow circle around it), timestamped "May 15, 2023, 2:07:31 PM", and status "scheduled disc".

27

Click "Create all tables"

s3://yusuf-cattaneo-bootcamp-nov2022/kaggle

cattaneo-bootcamp-nov2022/kaggle

Create all tables

Table name	Format	Changes	Results
vienna_weekends	CSV	New table	 Preview
vienna_weekends_refined-b694a4af...	ICEBERG	New table	 Preview
bird_feeder_hive	ORC	New table	 Preview