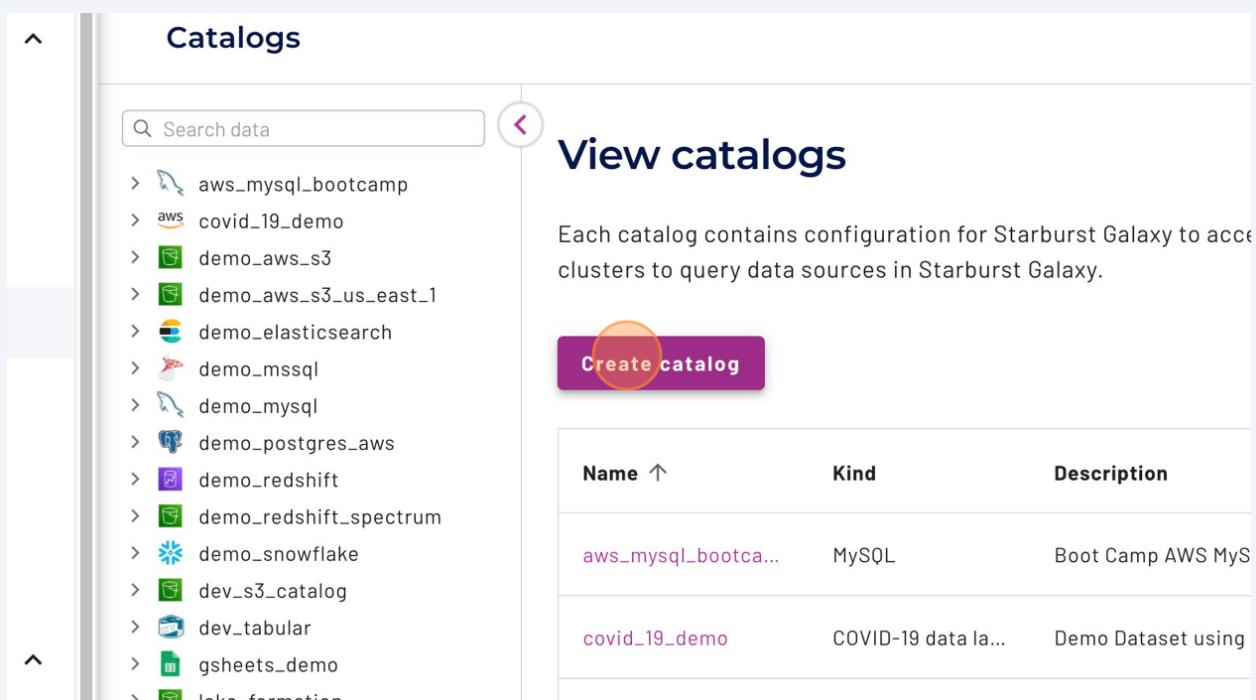


Creating a Catalog in Starburst Galaxy Configuration Hub

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 * ... ?

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 * ... ?

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 S3 bucket name *" field.

Metastore configuration

Configure access to the metastore to provide metadata and mapping information about the objects stored in Amazon S3. To change metastore type, create a new catalog.

Metastore type Starburst Galaxy

Default S3 bucket name *



Default directory name *



Allow creating external tables



Allow writing to external tables



Default table format

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

The screenshot shows the 'Metastore configuration' section. On the left, there's a sidebar with navigation items like 'Actions', 'Settings', and 'Pages'. The main area has a heading 'Metastore configuration' with a sub-section 'Metastore type: Starburst Galaxy'. It includes fields for 'Default S3 bucket name *' (set to 'starburst-galaxy-workshop') and 'Default directory name *' (highlighted with an orange circle). Below these are two toggle switches: 'Allow creating external tables' (unchecked) and 'Allow writing to external tables' (unchecked).

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

This screenshot is similar to the previous one but shows the 'Allow creating external tables' toggle switch being interacted with. The sidebar and top section are identical. In the main area, the 'Allow creating external tables' toggle switch is highlighted with an orange circle, indicating it is the target of the click action.

- 13** Click the "Allow writing to external tables" field.

Configure access to the metastore to provide metadata and mapping information about the objects stored in Amazon S3. To change metastore type, create a new catalog.

Metastore type Starburst Galaxy

Default S3 bucket name * starburst-galaxy-workshop

Default directory name * org_a/*

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 *

- 14** 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

15 Click "Connect catalog"

The screenshot shows the 'Test connection' step of the catalog creation process. At the top, there are three radio button options: 'Iceberg' (selected), 'Hive', and 'Delta Lake'. Below this is a section titled 'Test connection' with the sub-instruction: 'Validate that the network configuration allows Starburst Galaxy to connect to the data source.' Under 'Detected regions:', it lists 'aws US East (Ohio)'. A green success message box contains the text: 'Hooray! You can now add this catalog to a cluster.' At the bottom, there are two buttons: 'Test connection' (disabled) and 'Connect catalog' (highlighted with a red circle).

16 Click the "Read-only catalog" field.

The screenshot shows the 'Set permissions' step. On the left, a vertical list of steps is shown: 1. Select a data source (done), 2. Configure the connection (done), 3. Set permissions (optional) (highlighted with a red circle), and 4. Add to cluster (optional). To the right, the 'Set permissions' section is detailed. It starts with the heading 'Set permissions' and a note: 'Now that your **example_object_storage** catalog has been created, assign users access with roles. [Learn how to create roles here.](#)' Below this is the 'Catalog-level permissions' section, which includes a 'Read-only catalog' option (highlighted with a red circle) and a description: 'Prohibits all users, **including the catalog owner**, from modifying data or metadata in this catalog.' At the bottom is the 'Role-level permissions' section, which states: '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](#).'

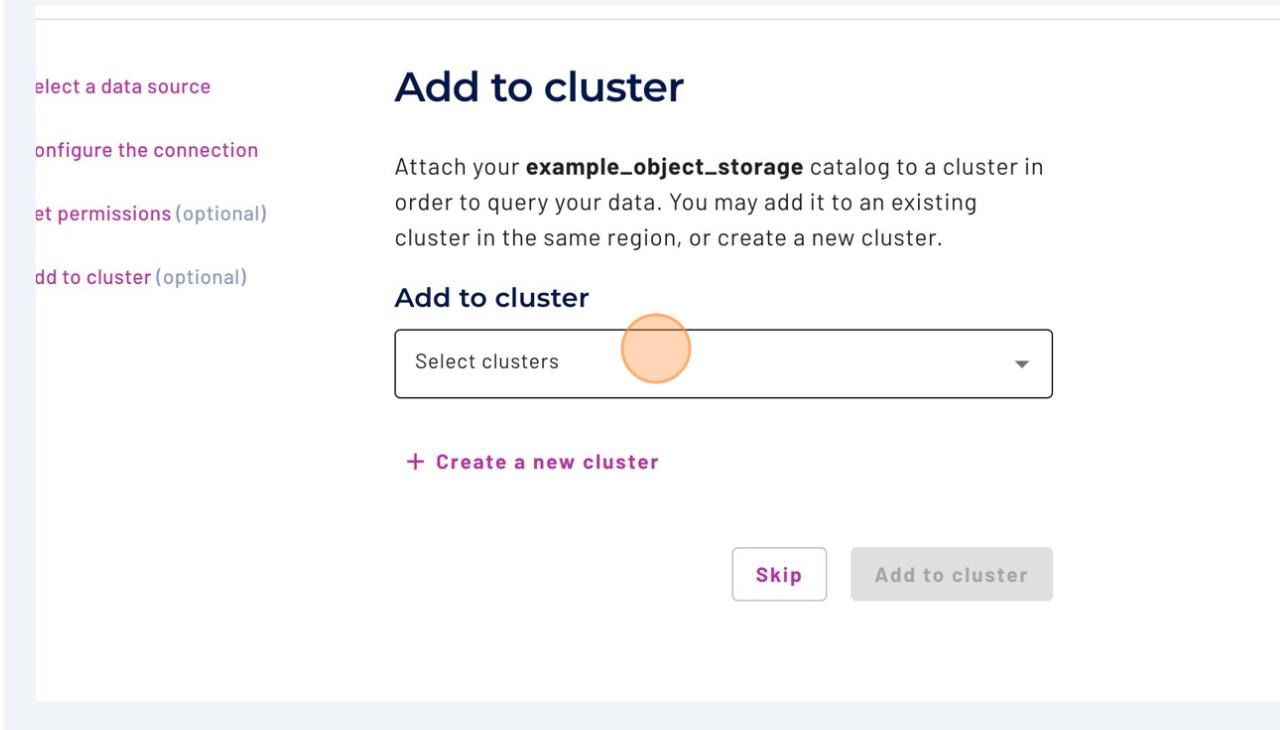
17 Click the "Read-only catalog" field.

The screenshot shows a step in a catalog creation wizard titled "Set permissions". On the left, there is a vertical list of steps: 1. Select a data source (checkmark), 2. Configure the connection (checkmark), 3. Set permissions (optional) (highlighted with a purple circle), and 4. Add to cluster (optional). The main content area is titled "Set permissions" and contains the following text: "Now that your **example_object_storage** catalog has been created, assign users access with roles. [Learn how to create roles here.](#)" Below this is a section titled "Catalog-level permissions" with a sub-section for "Read-only catalog": "Prohibits all users, **including the catalog owner**, from modifying data or metadata in this catalog." At the bottom, there is a section titled "Role-level permissions" with the text: "The following roles will be able to read data and metadata from all schemas and tables within this catalog, as described in [the documentation](#)." A dropdown menu labeled "Roles with read access" shows the role "accountadmin" selected.

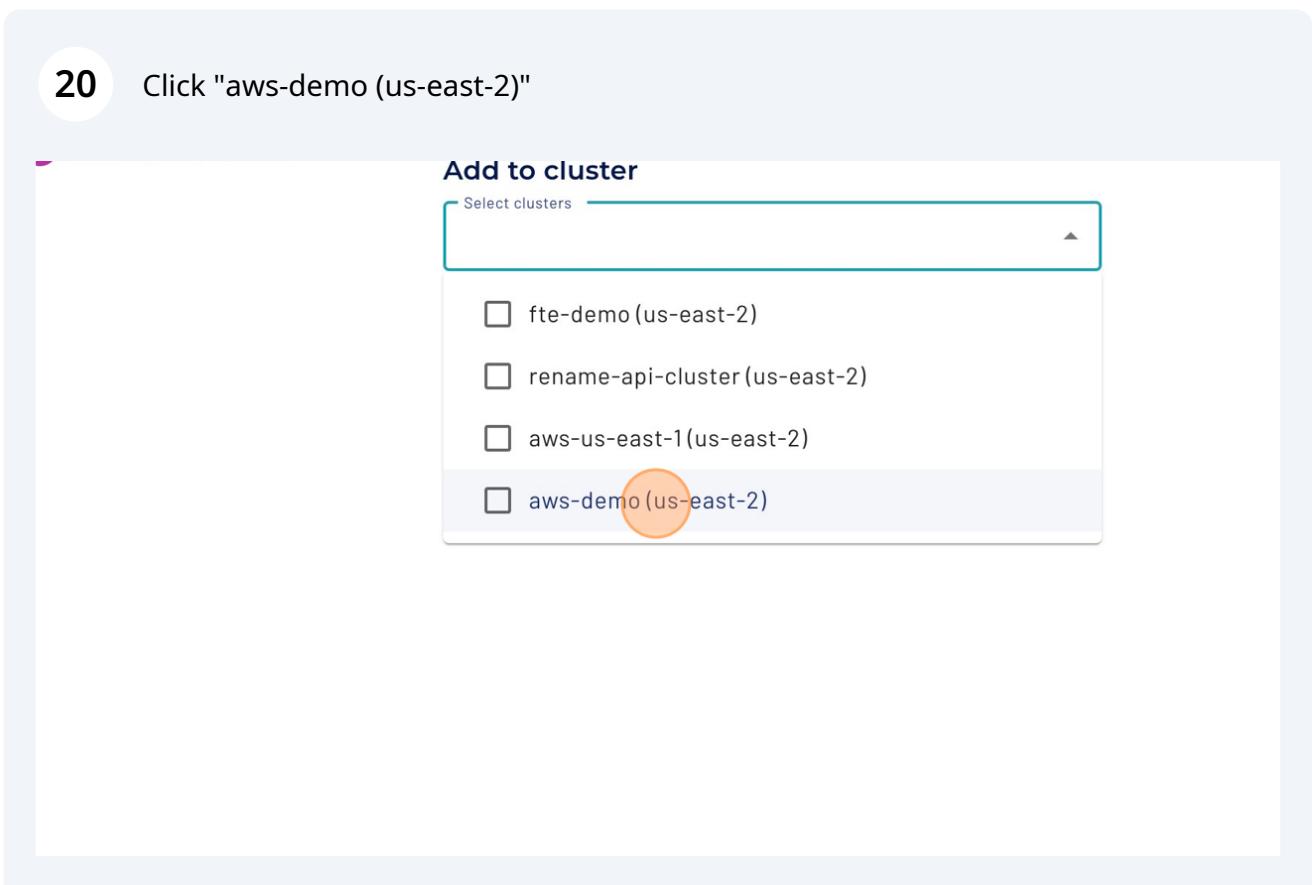
18 Click "Save access controls"

The screenshot shows a step in a catalog creation wizard titled "Save access controls". The main content area contains the text: "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](#)." Below this is a dropdown menu labeled "Roles with read and write access" showing the role "accountadmin" selected. Further down, another dropdown menu labeled "Roles with read access" also shows "accountadmin" selected. At the bottom right, there are two buttons: "Skip" and "Save access controls", with "Save access controls" being highlighted with a purple background and orange border.

- 19 Click the "Select clusters" field.



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



21 Click here.

- rename-api-cluster(us-east-2)
- aws-us-east-1(us-east-2)
- aws-demo (us-east-2)



22 Click "Add to cluster"

Add to cluster

Select clusters

aws-demo X



+ Create a new cluster

Skip

Add to cluster



23 Click "Run discovery"

Now that you have successfully configured your catalog, would you like to discover the schema within it to ensure that everything looks as expected?

All currently running queries will complete on the existing cluster and all new queries will be redirected to the new configuration as soon as it is ready.

[Do this later](#) [Run discovery](#)

24 Click "Run schema discovery"

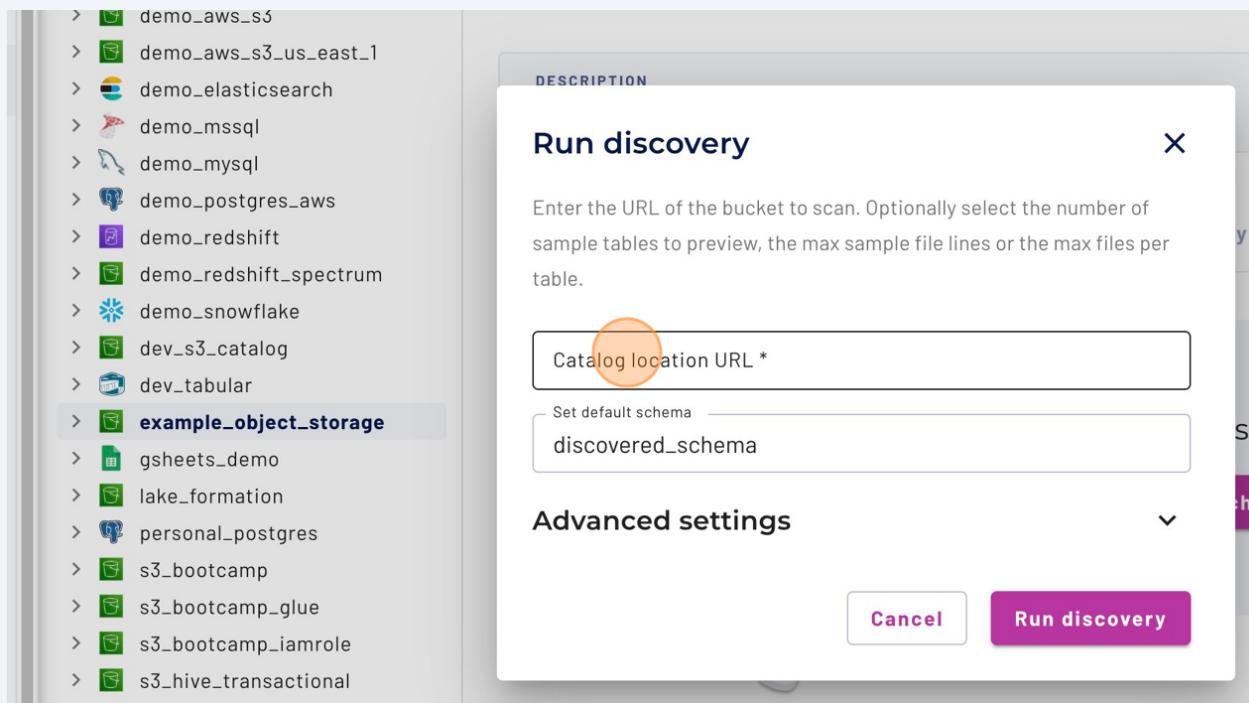
Schema discovery 0 Metrics Query history 0 Audit log Privileges



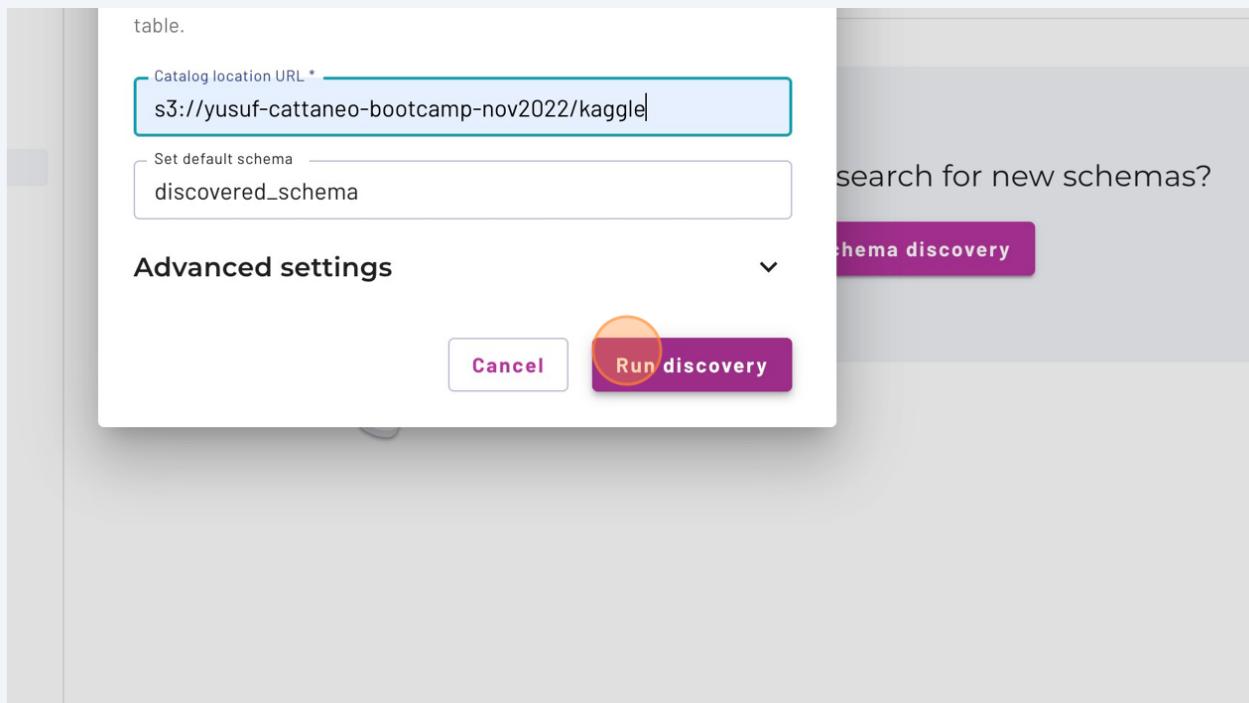
Would you like to search for new schemas?

[Run schema discovery](#)

25 Click the "Catalog location URL *" field.



26 Click "Run discovery"



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

The screenshot shows the Schema discovery tab selected in a UI. On the left, there's a sidebar with various database and storage options like postgres_aws, redshift, redshift_spectrum, snowflake, s3_catalog, and tabular. Below that is a section for 'multiple_object_storage' containing entries such as ets_demo, .formation, onal_postgres, bootcamp, bootcamp_glue, bootcamp_iamrole, live_transactional, sole, sole_data, .hive, and .no_datalake. The main area has tabs for Schemas, Schema discovery (which has a badge of 1), Metrics, and Query history (0). A 'Run discovery' button is present. A table lists a single item: 'Source' (s3://yusuf-cattaneo-bootcamp-nov2022/kaggle), 'Timestamp' (May 15, 2023, 2:07:31 PM), and 'Status' (scheduled discovery). A tooltip for the source URL is visible.

28 Click "Create all tables"

The screenshot shows a table with three rows. The first row contains the URL 's3://yusuf-cattaneo-bootcamp-nov2022/kaggle'. The second row is titled 'Cattaneo-bootcamp-nov2022/kaggle' and contains a 'Create all tables' button, which is highlighted with a yellow circle. The third row contains table names: 'vienna_weekends', 'vienna_weekends_refined-b694a4af...', and 'bird_feeder_hive', each with their respective formats (CSV, ICEBERG, ORC) and changes (New table). Each row also has a 'Preview' link.

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