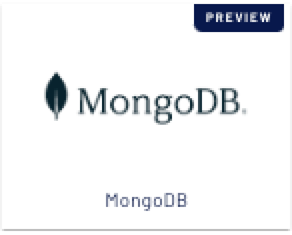
LAB 03 - Data Virtualization - Starburst

MongoDB catalogs

Use a MongoDB catalog to configure access to a MongoDB or MongoDB Atlas data platform.

Follow these steps to begin creating a catalog for MongoDB:

- In the navigation menu, select **Data**, then **Catalogs**.
- Click **Create catalog**.
- On the **Select a data source** pane, click the **MongoDB** icon.



- Follow the instructions in the next sections to configure your MongoDB connection.

Getting started with MongoDB

When configuring a MongoDB catalog in Starburst Galaxy, there are different access characteristics depending on whether your deployment uses a single cloud provider region or multiple cloud provider regions:

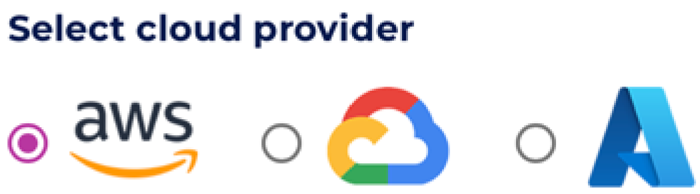
1. If your MongoDB instance uses a single node deployment or configures replica sets across the same cloud provider region, then the Starburst Galaxy catalog supports both read and write access.
2. If your MongoDB instance configures replica sets across multiple cloud provider regions, then the Galaxy catalog supports read-only access.

For a MongoDB Atlas federated database instance, the Starburst Galaxy catalog supports read-only access. Additionally, Starburst Galaxy supports tables with all rows conforming to same schema.

Select a cloud provider

The **Cloud provider** configuration is necessary to allow Starburst Galaxy to correctly match catalogs and clusters.

The data source configured in a catalog, and the cluster must operate in the same cloud provider and region for performance and cost reasons.



Select cloud provider

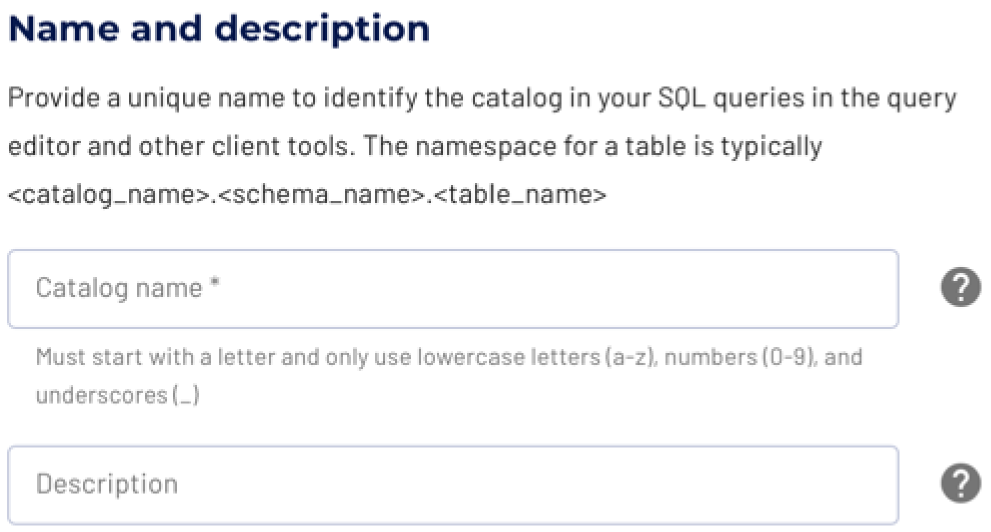
● aws ○ ○

Define catalog name and description

The **Catalog name** is visible in the query editor and other clients. It is used to identify the catalog when writing SQL or showing the catalog and its nested schemas and tables in client applications.

The name is displayed in the query editor, and in the output of a SHOW CATALOGS command. It is used to fully qualify the name of any table in SQL queries following the `catalogname.schemaname.tablename` syntax. For example, you can run the following query in the sample cluster without first setting the catalog or schema context: `SELECT * FROM tpch.sf1.nation;`.

The **Description** is a short, optional paragraph that provides further details about the catalog. It appears in the Starburst Galaxy user interface and can help other users determine what data can be accessed with the catalog.



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

Configure the connection

Read further to learn about each supported connection method. The following sections detail the setup for the supported cloud providers.

- **Connect directly**
  The connection to the database requires a username, password authentication, and the details necessary to connect to the database server, typically hostname or IP address and port.
- **Connect via SSH tunnel**
  A connection to the database can be established directly, if the Starburst Galaxy IP range/CIDR is allowed to connect.

  If the database is only accessible inside the virtual private cloud (VPC) of the cloud provider, you can use an SSH tunnel with a bastion host in the VPC.

- **PrivateLink** Starburst Galaxy supports AWS PrivateLink for MongoDB catalogs.

MongoDB configuration

To configure the connection to your MongoDB or Atlas data platform, provide the following details:

- **Database username:** specify a user with sufficient access rights to allow the desired querying. For example, if you only want to query data for analytics, the user does not need to have write access.
- **Database password:** specify the password for the user.
- **Hosts:** specify either a single host or multiple hosts if using MongoDB replica sets. For example, for a single MongoDB server with the connection string `mongodb://user:password@host1:port1`, the host value is `host1:port1`. For a MongoDB cluster using replica sets with the connection string `mongodb://user:password@host1:port1,host2:port2,host3:port3/`, the host value is `host1:port1,host2:port2,host3:port3`. If you do not provide a port, MongoDB uses port `27017` by default.
- **Use DNS seed list (SRV):** defaults to false. Enable this setting if you are using a DNS Seed List. The DNS seed list connection format begins with the prefix `mongodb+srv` instead of `mongodb`. In the connection string `mongodb+srv://cluster.example.mongodb.net/myFirstDatabase`, the host value is `cluster.example.mongodb.net`. Only a single host is allowed when using the `mongodb+srv` protocol.
- **Use TLS:** defaults to true. Enable this setting if you are connecting to a MongoDB instance with TLS.
- **Federated database:** defaults to false. Enable this setting if you are connecting to a MongoDB Atlas Federated Database. Use the connection string specific to your federated database instance. Contact your Starburst account team for support.

If you are using MongoDB Atlas and wish to connect directly, add the Starburst Galaxy IP Range/CIDR for your region to the Atlas IP Access List.

Note: If you are using the free tier of MongoDB Atlas, additional permissions are required for your databases to be queryable with Starburst Galaxy. You must create a custom role to allow actions on collections for your databases and collections to be viewable in Galaxy.

Test the connection

Once you have configured the connection details, click **Test connection** to confirm data access is working. If the test is successful, you can save the catalog.

If the test fails, look over your entries in the configuration fields, correct any errors, and try again. If the test continues to fail, Galaxy provides diagnostic information that you can use to fix the data source configuration in the cloud provider system.

Another way to resolve a connection failure is by temporarily opening traffic to the internet. To do this, add 0.0.0.0/0 to the MongoDB Atlas **IP Access List**. Once you receive a successful connection notification, delete 0.0.0.0/0 from the list.

Connect catalog

Click **Connect catalog**, and proceed to set permissions where you can grant access to certain roles.

Set permissions

This optional step allows you to configure read-only access or full read and write access to the catalog.

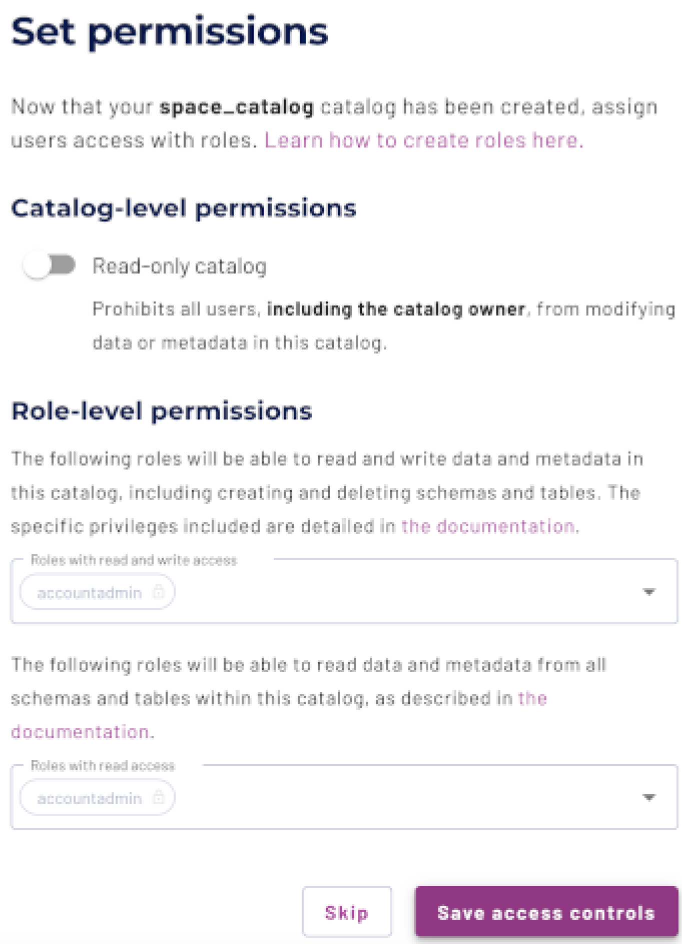
Use the following steps to assign read-only access to all roles:

1. Select the **Read-only catalog** switch to grant a set of roles read-only access to the catalog's schemas, tables, and views.
2. Next, use the drop-down menu in the **Role-level permissions** section to specify the roles that have read-only access.
3. Click **Save access controls**.

You can specify read-only access and read-write access separately for different sets of roles. That is, one set of roles can get full read and write access to all schemas, tables, and views in the catalog, while another set of roles gets read-only access.

Use the following steps to assign read/write access to some or all roles:

1. Leave the **Read-only catalog** switch cleared.
2. In the **Role-level permissions** section:
   - Expand the drop-down menu in the **Roles with read and write access** field and select one or more roles to grant read and write access to.
   - Expand the drop-down menu in the **Roles with read access** field and select one or more roles from the list to grant read-only access to.
3. Click **Save access controls**.



Set permissions

Now that your **space_catalog** 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.

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

Add to cluster

SQL support

The catalog provides read access and write access to data and metadata in MongoDB. It supports the following features:

- Globally available statements
- Read operations
- Write operations:

- Data management:
    - INSERT
    - DELETE
- Schema and table management:
    - CREATE TABLE
    - CREATE TABLE AS
    - DROP TABLE
    - ALTER TABLE
    - CREATE SCHEMA
    - DROP SCHEMA
    - COMMENT