Data Sharing and Collaboration

In this experiment, we will learn to use Data Sharing in Snowflake

Overview of data sharing at Snowflake

There are many ways to share data from your Snowflake account with users in other Snowflake accounts.

Why share data with Snowflake

When you use Snowflake to share data as a provider, you can manage who has access to your data, and avoid challenges keeping your data synchronized across different people and groups.

As a data consumer, you can reduce the data transformations you need to perform because the data stays in Snowflake, making it easy to join datasets shared with you with your own data.

If you share your data using listings, you can include metadata with your data share, such as a title and description, and usage examples to help consumers use the data quickly. In addition to the benefits for consumers, as a provider you get access to usage data, automatically replicate your data to other regions, and can even decide to charge for access to your data or offer some datasets publicly on the Snowflake Marketplace.

Options for sharing

Listings let you share data with people in any Snowflake region, across clouds, without performing manual replication tasks. If you use listings, you can provide additional metadata for the data that you share, view customer data usage, and for listings offered publicly on the Snowflake Marketplace, gauge consumer interest in your listings.

If you don’t want to share data using a listing, you can use a direct share instead. All methods of data sharing at Snowflake use Secure Data Sharing. No matter which option you choose, you can share with people who don’t have Snowflake accounts by using Reader Accounts.

Data Sharing Mechanism	Share With Whom?	Auto-fulfill Across Clouds?	Optionally Charge for Data?	Optionally Offer Data Publicly?	Get Consumer Usage Metrics?
Listing	One or more accounts in any region	Yes	Yes	Yes	Yes
Direct share	One or more accounts in your region	No	No	No	No

If you want to manage a group of accounts, and control who can publish and consume listings in that group, consider using a Data Exchange.

Listing

You can offer a listing privately to specific accounts, or publicly on the Snowflake Marketplace. For more about the Snowflake Marketplace, see About Snowflake Marketplace.

After you accept the provider and consumer terms, you can start sharing and consuming data shared with you with a listing. For more information, see About Listings.

Direct share

Use a Direct Share to share data with one or more accounts in the same Snowflake region. You don’t need to copy or move data shared with a Direct Share.

For more information, see Getting started with Secure Data Sharing.

Data Exchange

If creating listings that you offer privately to specific accounts isn’t an option, you can use a data exchange to share data with a selected group of accounts that you invite.

You must request that a data exchange be provisioned and configured for your account, then you can invite members to the exchange and specify whether they can consume data, provide data, or both.

The Data Exchange is supported for all Snowflake accounts hosted on non-VPS regions on all supported cloud platforms.

For more information, see Data Exchange.

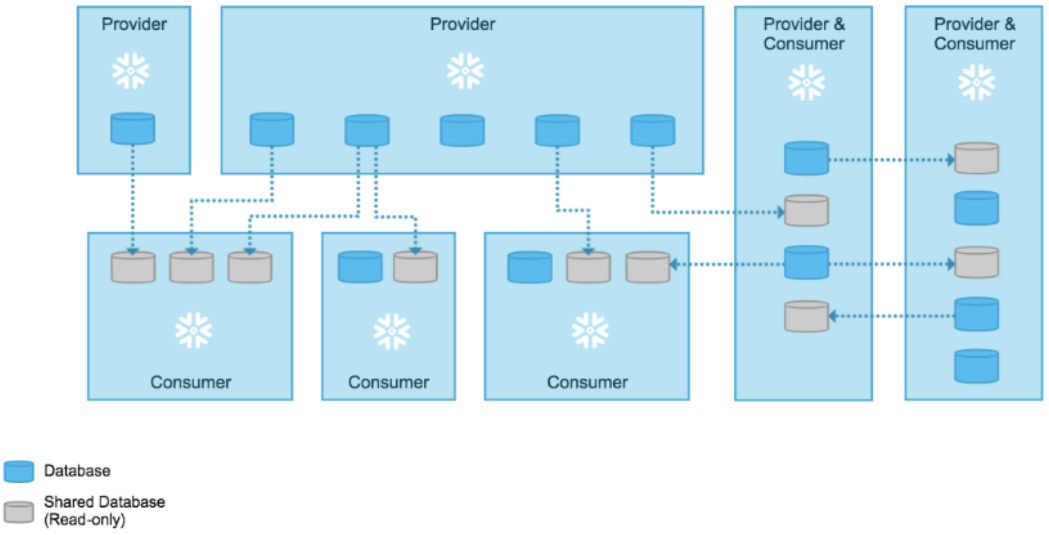
How does Secure Data Sharing work?

With Secure Data Sharing, ***no*** actual data is copied or transferred between accounts. All sharing uses Snowflake's services layer and metadata store. Shared data does not take up any storage in a consumer account and therefore does not contribute to the consumer's monthly data storage charges. The ***only*** charges to consumers are for the compute resources (i.e. virtual warehouses) used to query the shared data.

Because no data is copied or exchanged, Secure Data Sharing setup is quick and easy for providers and access to the shared data is near-instantaneous for consumers:

- The provider creates a share of a database in their account and grants access to specific objects in the database. The provider can also share data from multiple databases, as long as these databases belong to the same account. One or more accounts are then added to the share, which can include your own accounts (if you have multiple Snowflake accounts).
  For more details, refer to What is a share? (in this topic).
- On the consumer side, a ***read-only*** database is created from the share. Access to this database is configurable using the same, standard role-based access control that Snowflake provides for all objects in the system.

With this architecture, Snowflake enables a network of providers that can share data with multiple consumers (including within their own organization) and consumers that can access shared data from multiple providers:



1. **Create new Citibike_Pipeline database**

```
DROP DATABASE IF EXISTS CITIBIKE_PIPELINES;

-- Switch Context
USE ROLE ACCOUNTADMIN;

--Create the Warehouse
CREATE WAREHOUSE IF NOT EXISTS DATAPIPELINES_WH
    WITH WAREHOUSE_SIZE = 'XSMALL'
    AUTO_SUSPEND = 60
    AUTO_RESUME = TRUE;

--- Create the database and grant access to the new role create
CREATE DATABASE IF NOT EXISTS CITIBIKE_PIPELINES;

-- Switch Context
USE CITIBIKE_PIPELINES.PUBLIC;
USE WAREHOUSE DATAPIPELINES_WH;


-- Create the table for Trips
CREATE OR REPLACE TABLE TRIPS
(tripduration integer,
  starttime timestamp,
  stoptime timestamp,
  start_station_id integer,
  start_station_name string,
```

```
  start_station_latitude float,
  start_station_longitude float,
  end_station_id integer,
  end_station_name string,
  end_station_latitude float,
  end_station_longitude float,
  bikeid integer,
  membership_type string,
  usertype string,
  birth_year integer,
  gender integer);


-- Create the stage with the S3 bucket
CREATE or replace STAGE CITIBIKE_PIPELINES.PUBLIC.citibike_trips URL =
's3://snowflake-workshop-lab/citibike-trips-csv/';

list @citibike_trips;


-- Define the file format
create or replace FILE FORMAT CITIBIKE_PIPELINES.PUBLIC.CSV
    COMPRESSION = 'AUTO'
    FIELD_DELIMITER = ','
    RECORD_DELIMITER = '\n'
    SKIP_HEADER = 0
    FIELD_OPTIONALLY_ENCLOSED_BY = '\042'
    TRIM_SPACE = FALSE
    ERROR_ON_COLUMN_COUNT_MISMATCH = TRUE
    ESCAPE = 'NONE'
    ESCAPE_UNENCLOSED_FIELD = '\134'
    DATE_FORMAT = 'AUTO'
    TIMESTAMP_FORMAT = 'AUTO'
    NULL_IF = ('');


alter warehouse DATAPIPELINES_WH set WAREHOUSE_SIZE = 'LARGE';
copy into trips from @citibike_trips file_format=CSV pattern = '.*.*[.]csv[.]gz';
alter warehouse DATAPIPELINES_WH set WAREHOUSE_SIZE = 'XSMALL';
```

2.

3. **Let's create a database share**

This option will assign privileges from snowflake to the share.

```
USE ROLE accountadmin;

CREATE SHARE share_citi;

GRANT USAGE ON DATABASE CITIBIKE TO SHARE share_citi;
GRANT USAGE ON SCHEMA CITIBIKE.PUBLIC TO SHARE share_citi;
GRANT SELECT ON TABLE CITIBIKE.PUBLIC.TRIPS TO SHARE share_citi;

SHOW GRANTS TO SHARE share_citi;
```

	created_on	privilege	granted_on	name	granted_to
1	2023-09-18 05:02:03.548 -0700	USAGE	DATABASE	CITIBIKE	SHARE
2	2023-09-18 05:02:04.002 -0700	USAGE	SCHEMA	CITIBIKE.PUBLIC	SHARE
3	2023-09-18 05:02:04.426 -0700	SELECT	TABLE	CITIBIKE.PUBLIC.TRIPS	SHARE

```
ALTER SHARE share_citi ADD ACCOUNTS=<[account to share]>;

SHOW GRANTS OF SHARE share_citi;
```

This option will assign privileges to a share via roles

```
USE ROLE accountadmin;

CREATE ROLE share_user;

GRANT USAGE ON DATABASE CITIBIKE TO ROLE share_user;
GRANT USAGE ON SCHEMA CITIBIKE.PUBLIC TO ROLE share_user;
GRANT SELECT ON TABLE CITIBIKE.PUBLIC.TRIPS TO ROLE share_user;

SHOW GRANTS TO ROLE share_user;

CREATE SHARE share_citi;
GRANT USAGE ON ROLE share_user TO SHARE share_citi;

ALTER SHARE share_citi ADD ACCOUNTS=<[account to share]>;

SHOW GRANTS OF SHARE share_citi;
```