Enriching your analytics with data shared through Snowflake

WHITEPAPER

Analytics driven organizations are making third party data from partners and trusted vendors an integral part of their analytics strategy. Unfortunately, current methods for sharing data—such as EDI, FTP and even e-mail—require significant effort and duplicative storage. Even cloud file services require the copying and sending of files, an inherently manual, static and time consuming process.

Snowflake Data Sharing is a new way to share data that allows you to incorporate practically unlimited amounts of continuously updated data into your analytics without manual effort or overhead.

This document will help you to understand the process of utilizing data that your data provider has shared with you through Snowflake and disseminating the power of that data throughout your organization. You'll even see how to combine your existing enterprise data with the datasets that have been shared with you. But first, it makes sense to evaluate Snowflake Data Sharing to understand how it works and what it can provide to your organization.

WHAT IS DATA SHARING?

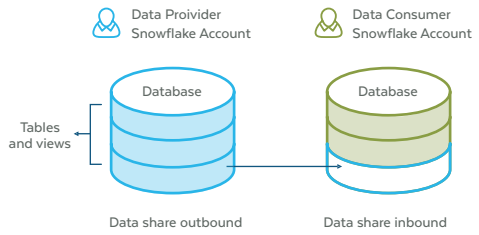
Snowflake Data Sharing makes it possible to directly share data in a secure, scalable way from the Snowflake **cloud data warehouse**. Organizations can use it to share data with other organizations with almost no friction or effort.

Existing approaches to data sharing (such as e-mail, FTP, EDI and APIs) have significant overhead that has prevented organizations from sharing data and consuming data effectively and efficiently. To start, the data would have to be extracted, copied and transferred by your data provider. The related schemas and metadata would need to be transported as well. Then, you would have to reconstruct and transform the data after transmission and configure your own data warehouse to store it. Perhaps most painfully, all of these efforts and tasks would need to be replicated every time the data is updated.

Snowflake is a cloud data warehouse that enables an entirely new way of sharing data that doesn't require the transmission of data, significantly reducing the traditional pain points of shared data.

Instead of transmitting data, Snowflake Data Sharing can be thought of as a way for you to directly access data that is in your data provider's account. Your data provider uses data sharing to give you the ability to view and query a dataset in their Snowflake instance.

Crucially, data sharing is only a component of the capabilities and storage within Snowflake itself. In the diagram below, you can see that the provider may share only a portion of their entire database with you, and on the same token you may have significantly more data within your Snowflake account that isn't related to the share. Because the data isn't copied or transferred between the accounts, the process of using and querying the data is instantaneous, and updates in the provider account are instantly available to you. What's more, data and queries are automatically optimized so there's no need for manual effort to use the data that's being shared with you.



GETTING STARTED WITH SNOWFLAKE

In order to begin consuming data that has been shared with you, you'll need to have a Snowflake account that is enabled for data sharing. This is actually quite straightforward, but you will need to contact Snowflake to **request an account** and have it enabled for data sharing. Since Snowflake is a data warehouse (that happens to also have the ability to share data), you can also use Snowflake with standard SQL to store, analyze and understand the other data that your organization uses.

Costs for data sharing

It's very important to note that there are no sign-up costs to create a Snowflake account. You pay Snowflake only for the compute capacity ("Snowflake Virtual Warehouses") you use to query data that has been shared with you. Crucially, you do not pay for the cost to store the data that is being shared with you, as this cost is borne by the provider. Your data provider may also charge you for the data they are sharing with you, but Snowflake does not participate in that transaction. If you choose to store your own data in Snowflake, then you will be charged as described in the **Snowflake pricing guide.**

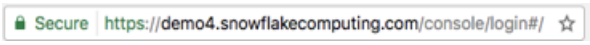
Logging into Snowflake and getting started

Once your Snowflake account has been created, it makes sense to begin with a quick introduction to the product itself. In order to use the data that has been shared with you, you'll need to understand how to create databases and warehouses and query your data with SQL.

There are detailed getting started tutorials on the Snowflake website that **can help you to begin quickly.**

ENABLING DATA SHARING AS A DATA CONSUMER

As noted before, you will need to have a Snowflake account enabled for data sharing before you begin. If you don't have an account, or it isn't enabled for data sharing, **contact Snowflake**. Once you have your account prepared and ready for data, be sure to send your account name to your data provider so that they can specify your account as a recipient of the data you need. To find your account name, simply look at the URL that you use to login to Snowflake; the account is the sub-domain that comes before "snowflakecomputing.com".



THE ACCOUNT IN THIS CASE IS "DEMO4"

ENABLING INBOUND SHARES AS A DATA CONSUMER

The process of consuming data through a data share is exceedingly simple. However, it is important to follow the data consumption process precisely.

1. Enable Snowflake inbound data sharing (**contact your Snowflake rep**).
2. View the Shares that have been shared with you.
3. Create a Database from the inbound Share.
4. Grant imported privileges to appropriate roles.

Once the account is enabled for data sharing, you can view any inbound data share by using the following SQL:

```
show shares;
```

The result in the Snowflake worksheet will show both inbound and outbound shares (if you have any), along with additional information on the database that is contained in the share, if any.

Results

row#	created_on	kind	name	database_name
1	2017-07-14 1...	INBOUND	DEMO22.STARTUP_SH...	STARTUP_DATABASE

The shares themselves are referred to with the following syntax:

```
[provider account].[share name]
```

In order to put the share to use, you need only create a database from the inbound share and apply the imported privileges to pertinent roles within your Snowflake account. The following syntax would enable anyone on your Snowflake account to view the data, but could be modified **to enable only specific roles.**

```
//Create a database from the share

create or replace database STARTUPS
from share DEMO22.STARTUP_SHARE;

//Query to view the data in the shared ta-
ble or view

select * from STARTUP_LIST;

//Grant permissions to others

grant imported privileges on database
STARTUPS to role public;
```

WHAT CAN YOU DO WITH YOUR INBOUND DATA SHARE?

At this point, the data will be visible in your account and can be queried with standard SQL. However, there are a couple of important differences between shared data and data that you have natively loaded into Snowflake.

1. You cannot alter the underlying data in a share. On the same token, when the data is updated in the provider account, the changes will immediately be visible in your account. In other words, the data is live and requires no transmission or copying for you to access as it updates.
2. If you want to alter a database object from an inbound data share, you may select and copy it into a new database. However, this new database will not be updated when data changes in the provider account.
3. You cannot share an inbound share with any other Snowflake account. In other words, you cannot share a share.

Otherwise, you are free to query, grant privileges, view and use the inbound data exactly as you would any other database in Snowflake. What's more, you can combine and utilize the inbound Data Shares with your existing corporate data in Snowflake without painful extract, load, transform (ELT) processes.

Data unification

Combining your new shared data with data that you've already loaded in Snowflake can be accomplished with straightforward standard SQL. Navigate to the database and schema you want to join to your shared data. Then, reference the shared database by using the following syntax:

```
[database name].[schema].[table]
```

The SQL for a simple join combining two separate databases would look like the following, assuming you have a current database that you are using in your existing session.

```
select * from STARTUPS.PUBLIC.STARTUP_LIST S
join CRUNCHBASE_LIST C on S.FIELD = C.FIELD;
```

Of course, you can change the SQL in whatever way you wish to execute the desired analysis. You can also join if you are currently using the shared database; in other words the direction or type of the join is irrelevant.

Business intelligence

Snowflake integrates with all of the leading Business Intelligence tools including Looker, Tableau, Microstrategy, Qlik and PowerBI. Most of these tools provide direct connections to Snowflake, which can enable you to share your new dataset throughout the organization efficiently by leveraging Snowflake's virtual warehouses to serve data to your BI tool quickly.

One enormous benefit to Snowflake Data Sharing is the ability to leverage the live connection for up-to-date business intelligence views. In other words, if the underlying database in the provider's account updates or changes, your BI dashboards and reports will update accordingly as they are refreshed.

Reference the **Snowflake resources webpage** to find more information about integrations with common BI tools.

CONCLUSION

Snowflake Data Sharing can help you augment the data that your organization is using with data from any (or many) Snowflake data providers. With low costs, and almost no friction, you can turn data that has been shared with you into valuable insights starting immediately and even power your existing analytics.

GET STARTED

To find out how Snowflake can help you get more value from your data, go to **https://www.snowflake.net/free-trial** and try Snowflake for free today.

Snowflake Computing, the cloud data warehousing company, has reinvented the data warehouse for the cloud and today's data. The Snowflake Elastic Data Warehouse is built from the cloud up with a patent-pending new architecture that delivers the power of data warehousing, the flexibility of big data platforms and the elasticity of the cloud at a fraction of the cost of traditional solutions. Snowflake is headquartered in Silicon Valley and can be found online at snowflake.net.