# Snowflake Data Sharing

EXTENDING THE BUILT-FOR-THE CLOUD DATA WAREHOUSE BEYOND ORGANIZATIONAL AND APPLICATION BOUNDARIES

WHITEPAPER

Data sharing is crucial to business operations. Retailers need to share sales data with vendors to manage inventory and supply chains. SaaS providers need to share the data they collect on behalf of their clients to support deeper customer and operational analytics. The list goes on.

To date, there has been no technology solution that organizations could turn to for sharing data. Traditional data warehouse platforms were not built to support the constant need to share data. They are too costly, inflexible and complex. As a result, organizations are forced to use a patchwork of solutions that include cumbersome methods such as FTP, APIs, email and file sharing.

Snowflake Data Sharing is a new innovation, available as part of Snowflake's data warehouse built for the cloud. Organizations can now externally share live data, at any scale, with other organizations while maintaining a single source of truth. Snowflake Data Sharing enables any organization to pursue new and imaginative ways to create insights and value from data.

## SNOWFLAKE DATA SHARING

Snowflake Data Sharing is a powerful yet simple-to-use feature of Snowflake for sharing data and for using shared data. In a matter of minutes, you can provide live access to any of your data stored in Snowflake for any number of data consumers, inside or outside your organization, without moving or copying the data. Share data across corporate divisions, external data consumers, and business partners to easily support richer analytics, new business models and data-driven initiatives.

With Snowflake Data Sharing, ready-to-use data is immediately available in real time. Query speeds are exponentially faster thanks to the limitless storage and compute resources of Snowflake's cloud-built architecture. Snowflake offers a new way to share data without the limitations and inefficiencies of existing solutions:

- Unlike file transfer approaches, such as FTP and email, Snowflake Data Sharing is far easier to use, provides instant access to live data and eliminates data copying or movement.
- Unlike cloud storage and file sharing services, Snowflake Data Sharing enables immediate querying of data in a secure, governed and controlled environment.
- Unlike electronic data exchange (EDI) and API-based approaches, Snowflake Data Sharing eliminates delays to viewing updated data, supports unlimited scale and allows unlimited concurrent access.

Fundamentally, traditional methods of data sharing address only one part of the challenge—moving data. Although traditional data warehouses and data lakes were designed to make data usable, they lack an architecture capable of meeting the needs of data sharing. Along with a lack of security and governance, among other things, their architectures cannot support concurrent access without cumbersome unloading and transfering in order to copy and move data from a data provider to their data consumers.

The lack of a comprehensive solution creates a struggle for data providers and consumers to easily share data. Cumbersome and incomplete data sharing processes also constrain the development of business opportunities from shared data.

## MADE POSSIBLE BY SNOWFLAKE'S BUILT-FOR-THE-CLOUD ARCHITECTURE

In contrast, Snowflake's patented multi-cluster, shared data architecture is the key to Snowflake Data Sharing. As a result, Snowflake's data warehouse as a service allows you to store, integrate and analyze all your data, share data, and use shared data, all from a single solution.

Snowflake Data Sharing is built on three key architectural innovations:

- Decoupling of storage and compute
- Global metadata and management
- Unlimited concurrency

### Independent Storage and Compute Scaling

The separation of storage and compute resources is a fundamental part of Snowflake's architecture. All data is stored, in optimized form, in the cloud built on Amazon S3. Data is managed by the Snowflake service, which capitalizes on the scalability, resiliency and near-infinite capacity of cloud storage. The data in cloud storage can be accessed concurrently by any number of independent compute clusters.

Decoupling of storage and compute is also critical for sharing data. It enables data consumers to directly access shared data. Unlike the monolithic architectures that bind storage and compute together, data in Snowflake can be accessed by any number of independent compute clusters without requiring multiple copies of the data. This unique architecture similarly provides all Snowflake customers the ability to share live data between them. Data providers can also make updates to data without contending for processing resources with other users, or from customers reading data at the same time.

### Global Metadata and Management

Making shared data usable requires access to data and coordination across all Snowflake customers and users to ensure consistency, security and performance. Snowflake's services layer is a key part of Snowflake's architecture. Global metadata, transactions and security are all managed from here, making it the control tower that tracks, logs and directs access to data for every database element and object contained within Snowflake.

The Snowflake services layer also provides critical services required for data sharing. It provides centrally managed control of access to data and ensures that data is secure at all times. Additionally, the services layer provides transactional consistency across all data providers and data consumers, ensuring that all data users see a consistent view of the data that it is always up to date. A data provider can update shared data in real-time. Likewise, all data consumers can view the data provider's updates and immediately query the shared data at the same time—all with transactional, ACID-based consistency.

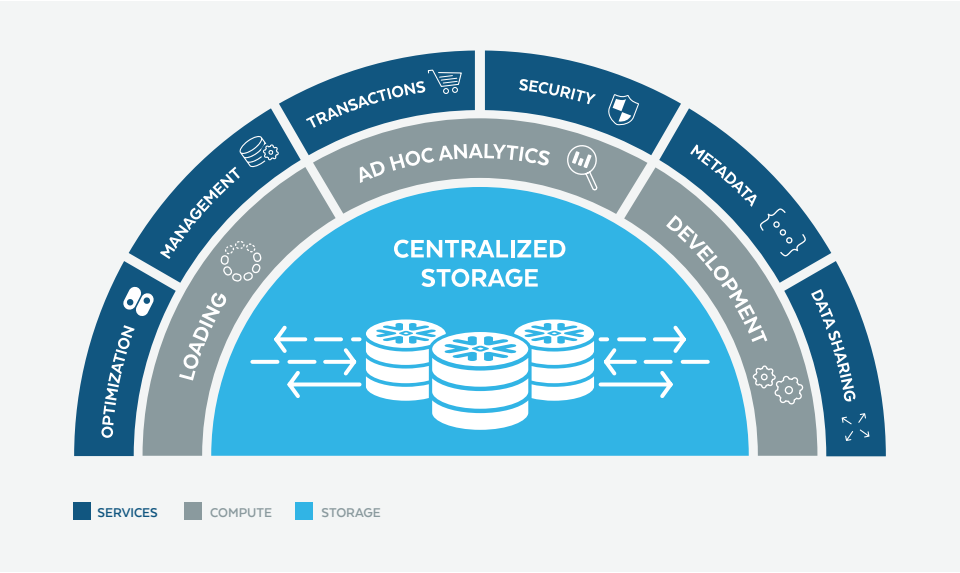


Fig. 1: The Snowflake Built for the Cloud Architecture – Separate Storage, Compute, and Services

### Unlimited Concurrency

With Snowflake, shared data can be accessed by large numbers of concurrent users and applications. In contrast, the architecture of traditional data warehouses forces all users to compete for resources, creating a struggle to deliver consistent performance. Snowflake's automatic concurrency scaling via multi-cluster warehouses takes simultaneous query processing even further, automating concurrency scaling within each Snowflake environment.

## USING SNOWFLAKE DATA SHARING

Snowflake Data Sharing allows sharing of a database and any objects contained within the database (schemas, tables, views, etc.) with any other Snowflake environment. When a database object is shared with a data consumer, the object remains in the data provider's Snowflake environment.

Data sharing is performed at the database level and all shared data are first-class objects. This means the shared data exists independently and can be manipulated and queried, along with any other database within a Snowflake environment. Within a shared database, Snowflake allows granular control of access to the objects through grants. Only objects granted access privileges are shared with other Snowflake users.

The data provider then GRANTS access to a data consumer. Instant access and no data copying or movement are made possible because all database objects are maintained and updated in Snowflake, and are orchestrated by Snowflake's global metadata management. Snowflake's global metadata management directs access to the shared data, according to the parameters established by the data provider via SQL semantics.

The data consumer, through their Snowflake environment, now has secure, read-only access to the database objects shared by the data provider. The data consumer can run analytics using whatever Snowflake resources are necessary from within their Snowflake environment. Organizations that do not already have their own Snowflake environment can quickly and easily sign up for the Snowflake service online and gain access to shared data through their new environment.

To share data, a provider pays only for the Snowflake storage and compute resources they use. The act of sharing data is at no cost. To query data, a consumer pays for only the Snowflake compute resources required to query the shared data. No storage costs apply for the data consumer unless they copy the data into a table.

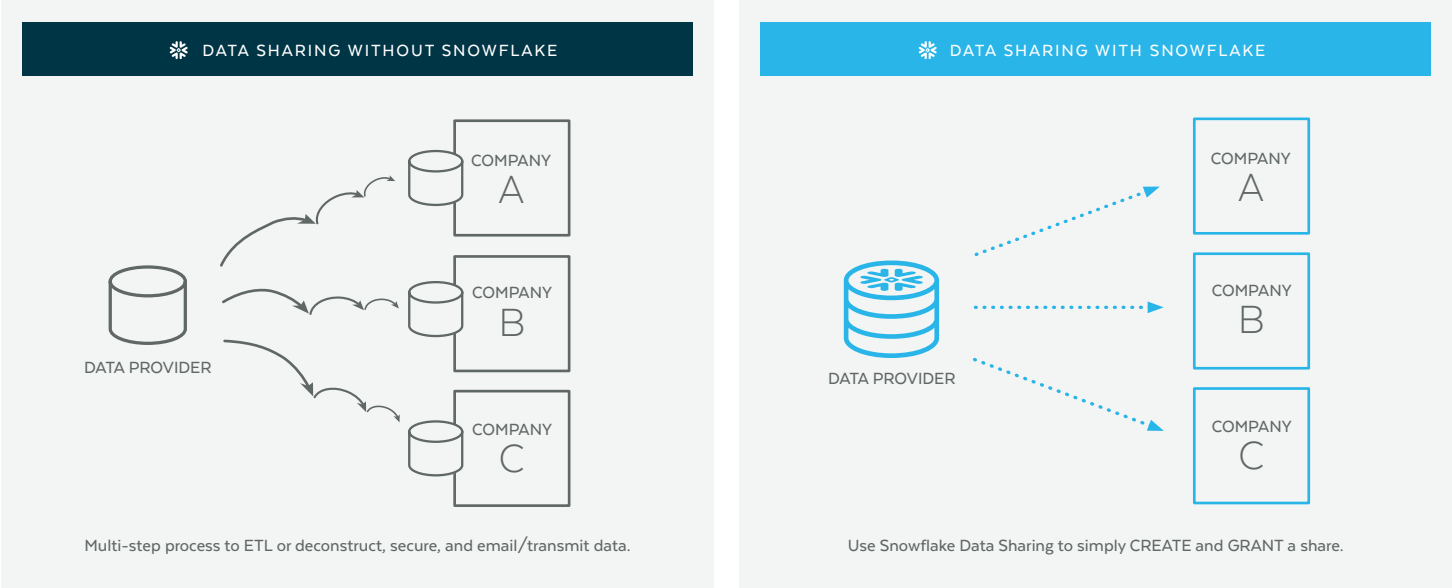


Fig. 2: Comparing alternatives—Snowflake makes data sharing easy

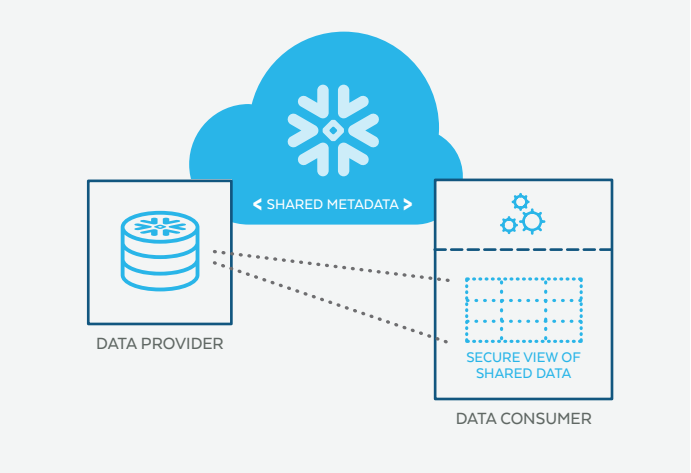


Fig. 3: How Snowflake Data Sharing Works

## UNLIMITED MULTI-TENANCY SCALABILITY

A critical capability of Snowflake's global metadata management is controlling access with secure views. Any number of data consumers can be granted access to the same database, but individual data consumers can view only the objects within the database for which they've been granted access.

With Snowflake, data providers have an easy method to manage multi-tenancy within a single database, as opposed to managing multiple separate databases

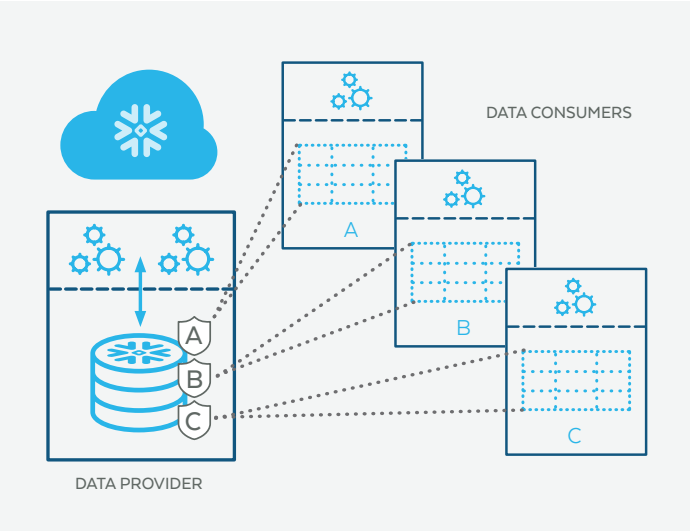


Fig. 4: Multi-tenancy Data Sharing with Controlled Access and Secure Views

for each data consumer. An example use case could be a sales CRM SaaS provider that maintains one master database for all CRM activities generated by its customers. When the sales CRM SaaS provider grants access to and shares data with its customers, the provider GRANTs data shares based on customer IDs (e.g. IDs = A, B, C), all from within one database. Simultaneously, the data provider can also execute queries on the shared database to support analytics within the data provider's business environment.

## ESTABLISHING DATA SHARING

The first step in sharing data is to specify what database objects to share with specified consumers. This is done via a data share object, effectively an "empty shell" that will house the references to the actual database and the shared database objects. Data shares are first-class objects in Snowflake for which Snowflake provides a set of DDL commands for creating and managing shares. Commands include CREATE SHARE, ALTER SHARE, DROP SHARE and others. Access commands include GRANT and REVOKE privileges.

Once a share is created, the data provider grants access to the specific database and database objects it shares. The SQL semantics are as follows:

### 1. Create the share

The following example creates an empty share named [sales_s]:

```
create share sales_s;
```

### 2. Add privileges for objects in the share

Grant usage on the primary object before granting usage on any objects within the primary object. For example, grant usage on a database before granting usage on any schemas contained within the database. Complete all grants for the data share before adding Snowflake data consumers.

The following example grants privileges for the [sales_db] database, the [aggregates_eula] schema and the [aggregate_1] table to the data-share object:

```
grant usage on database sales_db to
share sales_s;
grant usage on schema sales_
db.aggregates_eula to share sales_s;
grant select on table sales_
db.aggregates_eula.aggregate_1 to
share sales_s;
```

### 3. Confirm the contents of the share

```
show grants to share sales_s;
```

### 4. Share the database objects in the share object with the desired data consumers

The following example makes the [sales_s] share available to other Snowflake environments:

```
alter share sales_s add
accounts=data_consumerA, data_
consumerB;
```

Snowflake's environments, [data_consumerA] and [data_consumerB], are now able to see the share and create a database from it.

The above steps demonstrate that, with just a few simple commands, a data provider can easily share data with any number of data consumers.

## FROM DATA WAREHOUSE TO DATA SHAREHOUSE™

With unlimited data sharing and multi-tenancy capabilities, Snowflake extends the capabilities of the Snowflake built-for-the-cloud data warehouse to the Data Sharehouse. Snowflake Data Sharing enables organizations to easily forge one-to-one, one-to-many and many-to-many relationships to share data in new and imaginative ways.

### Industry Use Case Examples

Snowflake Data Sharing and the Data Sharehouse approach translate into a powerful but simple and cost-effective data warehouse for driving and expanding business intelligence and business assets from data. Industry examples include:

- **Adtech** – Share live pageviews, click stream, and more, directly with Adtech partners, driving more effective pricing for ad placements and faster responses for services.
- **Retail** – Share live sales data directly with vendors to assure the fastest possible and most accurate inventory and supply chain analytics and planning.
- **Gaming** – Share live gaming event data with developers, creative designers and other game production partners to enhance the gamer experience.
- **Healthcare** – Share read-to-query information, instantly, with medical groups and practices, hospitals, insurance companies and vendors to scale operations and reduce costs.

## CONCLUSION: SHARE AND IMAGINE MORE

From a data warehouse to a Data Sharehouse—with Snowflake Data Sharing, you can create a powerful, easy-to-use solution that enables you to share data with any number of organizations and sustain high levels of data processing concurrency, while maintaining a single source of truth. Organizations now have a new, never-before-available solution to share data, both internally and externally, that transforms how business assets are created from data.

## LET'S GET STARTED

Want to learn more about the benefits of Snowflake Data Sharing?
Visit our Data Sharing website.

Snowflake Computing is the only data warehouse built for the cloud. Snowflake delivers the performance, concurrency and simplicity needed to store and analyze all data available to an organization in one location. Snowflake's technology combines the power of data warehousing, the flexibility of big data platforms, the elasticity of the cloud and true data sharing at a fraction of the cost of traditional solutions. Snowflake: Your data, no limits. Find out more at snowflake.net.