DEFINITIVE GUIDE TO GOVERNANCE IN SNOWFLAKE

snowflake

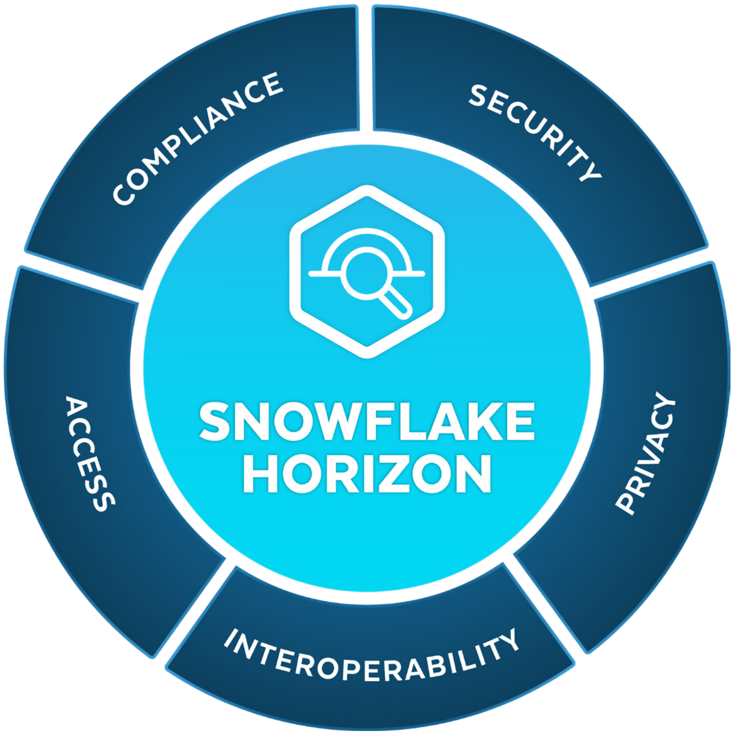
WHITE PAPER

TABLE OF CONTENTS

Introduction

Governance has never been more important for organizations. Data is proliferating across an increasing number of sources, applications and clouds, and in multiple formats. Generative AI and large language models (LLMs) are incorporating massive data sets that include sensitive data such as personal identifiable information (PII) and confidential medical and financial details. Companies must be able to unify, classify, protect, analyze and share this data to reach actionable insights while protecting it from unauthorized access and staying in compliance with regulations. According to an August 2023 Deloitte report, "Without robust data governance capabilities, the potential impact and value added by Generative AI will be severely limited and may even expose organizations to data and cybersecurity risks."

It's critical to have a strong governance foundation not only for data, but also for apps and AI assets, in one platform. Snowflake's single, cross-cloud governance model has always been a powerful differentiator, enabling customers to manage their increasingly complex data ecosystems with simplicity and ease. Now, Snowflake is enhancing its governance capabilities, which thousands of customers already rely on, through Snowflake Horizon.



Snowflake Horizon enables organizations to govern and discover their data, apps and more with a unified set of compliance, security, privacy, interoperability and access capabilities that is provided to customers without additional configurations or protocols. Here's how it works:

Compliance

With the number of regulations increasing around the world, companies that don't invest in compliance safeguards risk legal and financial consequences, reputational damage and operational disruptions. However, companies that develop mature compliance capabilities are better positioned to mitigate and forecast future risks, gain operational efficiencies and even grow their business. Snowflake Horizon includes a robust suite of compliance certifications, data quality monitoring, lineage, and cross-region, cross-cloud business continuity capabilities.

Security

Organizations face an endless torrent of security risks. They must get ahead of these risks by implementing processes and controls that block unauthorized access and misuse of sensitive content. Snowflake Horizon empowers you to secure your environment with continuous risk monitoring and protections, role-based access control (RBAC), and granular authorization policies that can be consistently applied across workloads, regions, and clouds.

Privacy

Organizations must deal with increasing privacy demands from consumers as well as regulators. This worldwide scrutiny makes effective privacy practices a crucial element of your organization's governance framework. Snowflake Horizon helps you unlock the value of your most sensitive data with a unique governance approach that preserves stringent data privacy standards. Snowflake Horizon enables advanced privacy policies such as aggregation, projection and differential privacy, along with Snowflake Data Clean Rooms, which makes it easy to build data clean rooms natively.

Interoperability

Governing content as it moves between different regions, clouds and platforms can be tricky, especially because you must adhere to the unique rules of each cloud provider. Integrating across a customer's entire data estate to maintain a unified view of governance demands significant time and resources. Snowflake Horizon addresses these challenges with a unique governance approach that values and prioritizes interoperability. With Snowflake Horizon, you can integrate with other Apache Iceberg-compatible catalogs and engines, as well as with key data catalog and governance partners.

Access

Accelerating access to content can lead to shorter project development cycles, deeper collaboration, better outcomes and more. But to reap these benefits, employees and partners must be able to find and share accurate, relevant content — and easily take action on it. With data estates constantly changing and growing, organizations face an ongoing battle to understand what's in all their content, and to grant appropriate access to it. Snowflake Horizon helps companies achieve these goals with a robust suite of governance capabilities that allows customers to classify, share, discover and immediately act on data, apps and more across regions and clouds.

This Definitive Guide to Governance in Snowflake will take a deep dive into how you can govern your environment and discover your content with Snowflake Horizon to build a trusted and secure data foundation that helps you accelerate success with gen AI and LLMs and any other use case.

COMPLIANCE

Today's organizations face a wide range of challenges — from economic pressures, shifting regulatory changes, and the rise of generative AI to accountability and environmental, social and governance (ESG) initiatives.

Leading organizations are moving away from viewing risk management as a perfunctory function and toward a more holistic, enterprise-wide approach built on the belief that risk appetite and risk tolerance can be used to help an organization achieve its strategic goals. In the 2023 Thomson Reuters Risk & Compliance Survey Report, 70% of respondents said they have noticed a shift over the past two to three years from check-the-box compliance to companies taking a more strategic approach.

Snowflake Horizon takes a two-pronged approach to helping customers address compliance requirements strategically:

- Compliance with **industry-standard data security and data privacy requirements**
- Compliance with **internal, self-imposed requirements** that protect sensitive data

Both goals involve similar requirements to protect data:

- **Audit** your data with reporting and monitoring, as well as lineage capabilities to gain visibility into what you are protecting.
- **Comply** with the relevant regulations and industry standards to protect data.
- **Establish business continuity** in the event of a disruption.

Companies that don't invest in compliance safeguards risk legal and financial consequences, reputational damage and operational disruptions. However, companies that develop mature compliance capabilities are better positioned to mitigate and forecast future risks, gain operational efficiencies and even grow their business.

Snowflake Horizon enables companies to achieve these benefits as it offers a robust suite of compliance certifications as well as powerful reporting and monitoring, lineage and business continuity capabilities.

Compliance certifications

Snowflake Horizon's certifications enable customers — especially those across the public sector — to secure their data on a platform with security and privacy protections.

Snowflake has received FedRAMP High Authorization on AWS GovCloud (US-Gov-West and US-Gov-East Regions). The Federal Risk and Authorization Management Program (FedRAMP) is a U.S. federal government-wide program that provides a standardized approach to security assessment, authorization and continuous monitoring of cloud products and services. This authorization enables the Snowflake platform to protect some of the federal government's most sensitive unclassified data across cloud computing environments.

Snowflake also recently achieved compliance with the UK's Cyber Essentials Plus (CE+), the FBI's Criminal Justice Information Services (CJIS) Security Policy, the IRS's Publication 1075 Tax Information Security Guidelines, and StateRAMP High, as well as a U.S. Department of Defense Impact Level 4 (DoD IL4) Provisional Authorization on AWS GovCloud and assessments by the Korea Financial Security Institute (K-FSI).

Snowflake is committed to meeting industry-standard regulatory compliance requirements. See a full list of Snowflake security and compliance reports.

Reporting and monitoring

To ensure they are meeting regulatory requirements and avoiding the consequences of noncompliance, organizations need to track and report on their data. Many regulations require organizations to maintain high-quality data, to monitor how and when data is accessed and changed, and to routinely audit their data to show the integrity and protection of sensitive and personal information. These practices are also important for achieving internal compliance and maintaining industry standards around data governance.

Snowflake Horizon offers robust data management and governance features that help organizations address these requirements. Here are some examples of its reporting and monitoring capabilities:

DATA GOVERNANCE INTERFACE

To make it easier for customers to manage and assign tags and policies, Snowflake Horizon offers an intuitive interface that provides an at-a-glance summary of the total objects, tagged objects and protected objects in an account. Customers can further drill down into the interface to generate customized reports with filters for Database/Schema, Tag and Object Kind, and even take instant action into a specific table, view or column to apply tags and corresponding data policies.

ACCOUNT USAGE VIEWS

This schema enables the querying of object metadata, as well as historical usage data, for the owner's account and all reader accounts. The views in this schema display object metadata and usage metrics for the account.

ACCESS HISTORY (READS)

Within the ACCOUNT_USAGE schema, this view provides information on when the user's query reads data, thereby providing visibility into column-level access to help customers understand usage. They can audit a log of tables, views and columns accessed by each query to facilitate regulatory audits, and provide insights on popular and frequently accessed tables and columns.

SCHEMA CHANGE TRACKING

This advanced feature enables tracking of changes to database objects such as tables, views and stored procedures.

DATA QUALITY MONITORING

To effectively monitor and report on data quality degradation across their organization, customers can use Data Quality Monitoring (in private preview) to either access out-of-the-box system metrics or create custom metrics. They can also define the frequency for automatically measuring the quality of their data and configure how to receive email notifications when quality thresholds are violated.

Lineage

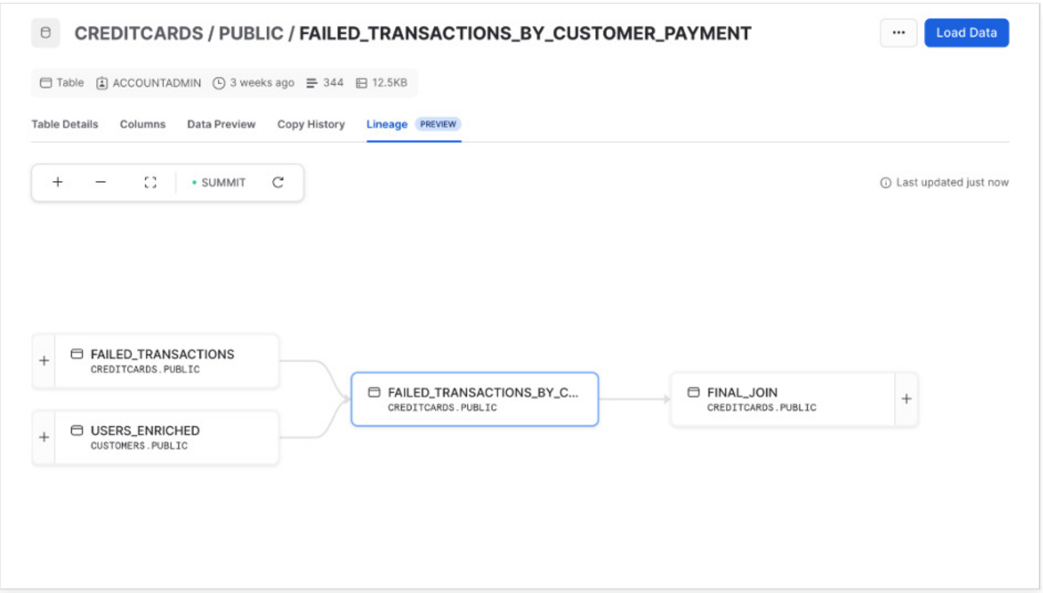
Regulations and or internal policies require that organizations protect their sensitive and personal data. But with massive volumes of data flowing through multiple companies' borders, it becomes increasingly challenging to track and secure it all. Snowflake Horizon's lineage capabilities help track the flow of data over time, providing a clear understanding of where the data originated, how it has changed, who has rights to it, and who has accessed or copied it. Here are several features that allow users to track data lineage:

ACCESS HISTORY (WRITES)

This view provides information on when a SQL statement performs a data write operation, such as INSERT, UPDATE and DELETE, along with variations of the COPY command, from the source data object to the target data object. Snowflake Horizon provides tracking of all the write operations to effectively provide a view of data lineage.

OBJECT DEPENDENCIES

Object dependencies provide a way to track the relationship between different data objects, which is crucial for maintaining data lineage. Object dependencies provide details on how data is generated, captured, modified and used by triggering dependencies based on the name and/or ID of the object.



SNOWFLAKE'S LINEAGE UI PROVIDES A VIEW OF UPSTREAM AND DOWNSTREAM OBJECT LINEAGE.

LINEAGE UI

The Lineage UI (currently in private preview) gives customers a bird's-eye view of the upstream and downstream lineage of objects. With this UI, customers can easily visualize how downstream objects may be impacted by modifications that happen upstream. In addition, data governors can take bulk actions to propagate tags and policies to protect all downstream columns that have personally identifiable information.

Business continuity

Prioritizing business continuity empowers organizations to uphold compliance standards. In the event of a massive outage (due to network issues, a software bug, etc.) that disrupts cloud services in one region or across that entire cloud service provider (CSP), organizations must be able to maintain or quickly resume operations. To enable continued availability and data durability in such a scenario, Snowflake Horizon allows customers to replicate their critical account objects to another Snowflake account in their organization in a different region and/or cloud. Snowflake Horizon's single, cross-cloud governance model enables customers to manage and maintain their data environments across regions and clouds, providing a robust framework for ensuring business continuity.

RECOVERY POINT OBJECTIVE AND RECOVERY TIME OBJECTIVE

There are two critical metrics in disaster recovery and business continuity planning:

- Recovery point objective (RPO) refers to the state of the data at the time of system recovery and determines the level of data loss acceptable to the business. This becomes a function of how frequently changes are copied or replicated to an alternate location. The time that passes between the last replication operation and a disaster event is the RPO.
- Recovery time objective (RTO) is the maximum amount of time an organization will tolerate an application not being available in the event of a disaster.

Snowflake Horizon's business continuity capabilities, guided by RTO and RPO, help organizations to easily safeguard mission-critical accounts and data sets to maintain uptime and address compliance requirements.

REPLICATION AND FAILOVER/FAILBACK

This feature enables account metadata — including everything from user identity- and role-based access controls to governance policies, warehouses and resource monitors — to be automatically synchronized across clouds and regions for continuous availability. Customers can replicate and fail over databases, shares, role-based access control, governance policies, compute resources, network policies and more.

Replication uses two Snowflake objects, **replication group** and **failover group**, to replicate a group of objects with point-in-time consistency from a source account to one or more target accounts. A replication group allows customers to specify what to replicate, where to replicate to, and how often. This means specifying, at customizable scheduled intervals, which objects to replicate and to which regions or cloud platforms. A failover group enables the replication and failover of the account objects in a group.

CLIENT REDIRECT

This feature allows for the seamless redirection of client connections across Snowflake accounts in different regions and clouds. In the event of an outage, Snowflake Client Redirect facilitates seamless failover from primary to secondary so apps and users can continue functioning without disruption. When the outage is resolved, the organization can redirect client connections back to the original primary connection.

PFIZER FOCUSES ON STRATEGIC GROWTH

In its quest to digitize drug discovery and manufacturing, pharmaceutical company Pfizer uses Snowflake to improve patient outcomes, reimagine clinical trials through robotics and automation, and perform predictive analytics for diagnosis and supply chain tracking.

As a global company with teams located across the Americas, Europe and Asia, Pfizer views data as playing a critical role in every aspect of its operations, from research and development to manufacturing and distribution. The company's success hinges upon having reliable access to the same data at all times, and the ability to seamlessly share data sets within its different business units.

Pfizer used Snowgrid — Snowflake's cross-region, cross-cloud technology layer that powers business continuity in Snowflake Horizon — to safeguard its most important data sets from potential cloud-specific outages. Each business unit at Pfizer owns various data sets, and each set has varying levels of importance and user-based access clearance. Snowflake's highly configurable replication features allow each team to choose the replication frequency that best suits each specific use case. This flexibility ensures that Pfizer can optimize the replication process to meet the RPO and RTO requirements of each business unit. Across Pfizer, Snowflake ensures business continuity for over one petabyte of data and hundreds of applications, which can be seamlessly redirected near instantly in the event of an outage.

This consistency reduces the operational burden on Pfizer's data teams so they can concentrate on their data-intensive projects instead of wasting time on replication maintenance. Today, Pfizer can focus on new strategic projects rather than worrying about the underlying cloud data platforms.

Read more about Pfizer's Snowflake story.

SECURITY

Organizations face an endless barrage of security and privacy risks. They must get ahead of these risks by implementing processes and controls that prevent unauthorized access and misuse of sensitive content including data, apps and more. A successful security strategy protects organizations from financial loss and legal liability, and helps safeguard their business reputation.

Organizations need to draft comprehensive governance policies that define who has access to sensitive content and provide tools to track any potential misuse. In addition, organizations must incorporate access controls that regulate individual access to sensitive content. These controls can be specific, granting access rights to individual records or files if necessary.

Snowflake Horizon helps organizations grant access to the content that different teams need while also protecting it. This allows teams to access what they need while preserving the organization's security framework in place that instills confidence in the reliability and safety of the services, and the comfort of knowing the platform will be secure even in the face of major cyberattacks.

Snowflake Horizon takes a two-pronged approach to helping customers address compliance requirements strategically:

- Continuous risk monitoring and protections
- Role-based access control (RBAC)
- Granular authorization policies

Snowflake Horizon enables organizations to go above the baseline security requirements with their governance practices across platform security, account security, data security, and apps and model security. Let's take a closer look at each of these.

Platform security

Platform security entails putting in place a security foundation that is made up of tools, processes and architecture. It specifically combines software and hardware to secure IT infrastructure, network components, storage, and the operating systems and applications that live on those platforms.

Threat hunting is one powerful technique to stay ahead of threat actors. By actively searching for potential threats, Snowflake Horizon helps organizations identify threats that may have evaded standard security measures, including firewalls and antivirus software. Threat hunting helps organizations quickly identify potential attacks and create effective strategies to stop them before they are exploited by attackers. Additionally, Snowflake has a sophisticated incident response team aligned with widely adopted standards of operation, comprehensive vulnerability management systems, and regular third-party audits, pen tests, and bug bounty programs for independent researchers and security engineers to report bugs. Furthermore, dedicated offensive security professionals who are experts in attacking systems regularly conduct tests on Snowflake's platform security capabilities.

Snowflake provides an integrated approach to platform security with controls, combined with automated and continuous compliance monitoring systems, that meet sophisticated and complex security requirements. The Snowflake Shared Responsibility Model is built on an ethos to minimize the customer's side of the Shared Responsibility Model through automation and feature availability within Snowflake's product offering.

The model gives organizations a more comprehensive view of their security obligations as well as the measures Snowflake has implemented to protect the platform. The model creates transparency and allows organizations to actively participate in securing their Snowflake deployments.

There are four key benefits to the Snowflake Shared Responsibility Model:

1. **Clear accountability:** The model clearly defines the responsibilities between Snowflake and its customer, eliminating any ambiguity and fostering a collaborative approach to security.
2. **Enhanced security:** The model enables both the platform provider and users to actively contribute to the security of Snowflake deployments, creating a robust security posture. For example, the Snowflake platform allows organizations to restrict access to a Snowflake account based on source IP addresses. They can configure the list of trusted IP addresses by setting up account-level network policies.
3. **Improved compliance:** Adhering to the Shared Responsibility Model helps organizations demonstrate their commitment to security and compliance requirements.
4. **Ease of use:** Security is built into Snowflake services, and the secure defaults available within the Data Cloud help minimize the customer's maintenance burden for their obligations under the Shared Responsibility Model as much as possible.

Account security

Snowflake expands upon the Shared Responsibility Model by creating a benchmark that captures the Data Cloud's security capabilities and security best practices in partnership with the Center for Internet Security (CIS). This CIS Snowflake Foundations Benchmark is a set of industry-recognized best practices and security configurations to help organizations confirm if their account is conforming to high security standards even if they are not security experts.

To help customers better discover security risks while providing recommendations to resolve these issues, Snowflake launched the Trust Center (currently in private preview). The Trust Center streamlines cross-cloud security monitoring in one centralized place to reduce security monitoring costs, resulting in lower total cost of ownership (TCO) and the prevention of account risk escalations. The Trust Center provides a programmatic way to verify that both Snowflake and customers are fulfilling their respective responsibilities in the Snowflake Shared Responsibility Model to uphold industry best practices outlined in the CIS Snowflake Foundations Benchmark.

Data, apps and model security

Data, apps and model security refers to how organizations protect data, applications and models from unauthorized access. Snowflake Horizon addresses this by combining two approaches:

- **Discretionary access control:** Each object has an owner, who can in turn grant access to that object.
- **Role-based access control:** Access privileges are assigned to roles, which are in turn assigned to users. Users can apply unified role-based access controls across data, models and Streamlit apps in Snowflake. The Snowpark Model Registry (in public preview) allows customers to securely manage models and their metadata in Snowflake, regardless of origin, with full role-based access control.

Snowflake Horizon also provides industry-leading features that enable high levels of security for your account and users, as well as all the data you store and access in Snowflake.

- **Row-level security:** Allows the application of a row access policy to a table or view to determine which rows are visible in the query result.
- **Masking policies:** his helps protect sensitive data from unauthorized access while allowing authorized users to access sensitive data at query runtime. This means that sensitive data in Snowflake is not modified in an existing table (that is, no static masking). Rather, when users execute a query in which a masking policy applies, the masking policy conditions determine whether unauthorized users see masked, partially masked, obfuscated or tokenized data.

- **Tag-based masking policies:** Protects column data by assigning a masking policy to a tag and then setting the tag on a database object or the Snowflake account.
- **Dynamic Data Masking:** This is a column-level security feature that uses masking policies to selectively mask plain-text data in the table and view columns at query time.
- **Conditional masking:** This uses a masking policy to selectively protect the column data in a table or view based on the values in one or more different columns.
- **External Tokenization:** Enables accounts to tokenize data before loading it into Snowflake and detokenize the data at query runtime. Tokenization is the process of removing sensitive data by replacing it with an undecipherable token. External Tokenization makes use of masking policies with external functions.

ABB UNIFIES DATA TO SAVE MILLIONS

With Snowflake, ABB built a scalable foundation for enabling a "data-first" mentality across its four business areas. Snowflake's role-based access controls and data governance features are empowering more users to explore data and develop solutions that accelerate innovation at ABB. "Instead of just providing access to a dashboard, we can share information that's specific to each business analyst and let analysts utilize models inside their solutions," says Michael Thorne, Global Analytics Product Engineering and Delivery Manager at ABB. "It's opened up greater opportunities for reuse."

Snowflake Marketplace offers a convenient, secure way to access data from ABB's data providers. For example, ABB's electrification business uses Snowflake Marketplace to access construction, real estate and commodity data products. According to Thorne, "Data is readily available, you don't have to build ingestion pipelines and insights are live as soon as the provider updates the information. Time to actionable insight is drastically reduced."

Read more about **ABB's Snowflake story**.

PRIVACY

Today's organizations face increasing privacy demands from consumers as well as regulators. Roughly 4 in 10 Americans say they are very worried about companies selling their information to others without them knowing (42%) or people stealing their identity or personal information (38%), according to a 2023 Pew Research study. Meanwhile, the number of privacy regulations worldwide continues to grow — already, 71% of countries have enacted data privacy laws, according to the United Nations.

The worldwide scrutiny of privacy makes effective privacy practices a crucial element of your organization's governance framework. They ensure that data is used effectively while being protected appropriately. They help businesses maintain trust with customers, comply with regulations, and avoid costly fines and reputational damage.

At the same time, sharing and analyzing data opens up untapped opportunities for organizations in every industry. These include:

- Leveraging AI and machine learning algorithms to uncover hidden insights in data
- Uncovering new revenue streams by monetizing first-party data
- Improving business processes through data-driven optimization
- Driving strategic decision-making based on factual evidence rather than intuition

But data sets often include sensitive data such as personal identifiable information (PII), proprietary data, and confidential medical and financial details. To share and analyze this type of information, data providers often must remove information such as identifying fields and transaction-level granularity, devaluing the data. Often, providers cannot share the data at all due to privacy concerns and regulatory restrictions.

Snowflake Horizon helps you unlock the value of your sensitive data with a unique governance approach that upholds stringent data privacy standards. This approach includes two powerful capabilities:

- Snowflake Data Clean Rooms, generally available in AWS East/West and Azure West, which enable you to easily collaborate on sensitive data with teams inside and outside your organization
- Differential Privacy, which allows you to share and protect sensitive data while retaining its analytical value

With the combination of Snowflake Data Clean Rooms and Differential Privacy, Snowflake Horizon offers a uniquely comprehensive, robust governance solution that ensures privacy and compliance for customers. Moreover, Snowflake's privacy capabilities are built-in and accessible through an easy user interface, so you don't have to be a privacy expert or programmer to protect your data.

Snowflake Data Clean Rooms

A data clean room is a secure environment that allows multiple parties to safely combine first-party data, ensuring compliance with privacy and regulatory requirements. Data clean rooms arose in response to stricter privacy and security laws around data and the ensuing changes from browser vendors. Advertisers, in particular, have used clean rooms for improved audience overlap analysis, better targeting, and accurate measurement and attribution. Today, they are a necessity for any organization that needs to collaborate on sensitive data without compromising its confidentiality, integrity or security.

Snowflake Data Clean Rooms is a native solution that provides governed access to shared data and allows multiple parties to collaborate on first-party data while preventing drill-down to sensitive information. What's unique about Snowflake Data Clean Rooms is that any developer or business user can quickly build a clean room through an easy-to-use interface, built on the Snowflake Native App Framework. This approach allows parties to collaborate in one environment without having to move data, letting each participant retain control and security of their information. And because of Snowflake's cross-cloud capabilities, organizations can connect to the clean room and ensure that the data is governed, regardless of their cloud region.

Here are some examples of how Snowflake Data Clean Rooms can be used to create data clean rooms natively in Snowflake to maintain privacy in collaboration across various industries:

- **Advertising, media and entertainment:** Advertisers can use data clean rooms to link first-party, customer marketing and advertising data from multiple parties for attribution. A media company can use a data clean room to share anonymized audience data with an advertiser for targeting. The advertiser can analyze this data to identify patterns and trends in audience behavior, and tailor their advertising campaigns accordingly without having access to identifiable information.
- **Healthcare:** A hospital can use data clean rooms to share highly sensitive patient data with a pharmaceutical company. The company can analyze this data to identify patterns in patient outcomes related to a specific drug without being able to identify or extract patient information. This enables privacy of the patients' data while still facilitating meaningful insights, which may ultimately lead to improved medical outcomes.
- **Retail:** Retailers can use data clean rooms to collaborate with brands that advertise with them. For example, a retailer can share transaction data in a privacy- and governance-safe manner to provide insights into conversion signals and achieve better targeting, personalization and attribution.
- **Financial services:** Asset managers at large banks can leverage customer data to understand trends in sector-level consumption and spending. But strict privacy regulations prevent the sharing of certain PII within banks and other financial institutions. A data clean room enables bankers to share their data in a way that protects the sensitive information and customer identities, while retaining the value of the data for asset management.
- **Technology:** Global technology companies often must share information between regions that have different regulations. For example, a company may have a product development team in the U.S. that collects user data for product improvement and a marketing team in the EU that could benefit from this data for market analysis and strategy development. The company could use a data clean room to securely share anonymized user data with the marketing team in the EU. The marketing team could then analyze this data to gain insights and develop strategies without having access to identifiable user information.

Different collaboration scenarios around sensitive data require different privacy-preserving solutions. Snowflake's robust suite of privacy-preserving controls and policy features can be used in combination with Snowflake Data Clean Rooms or independently outside of a data clean room. These advanced privacy policies include:

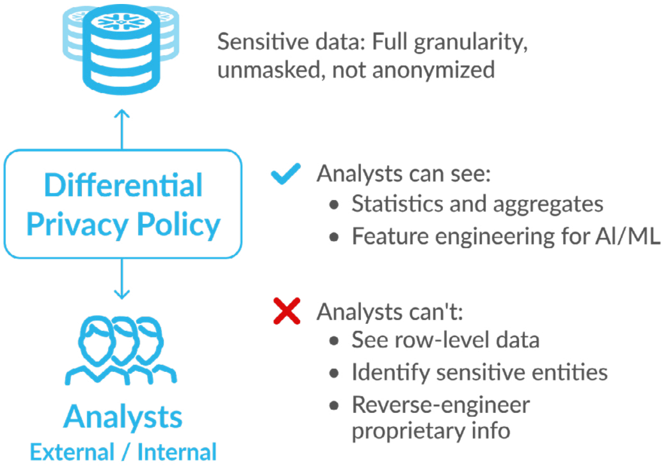
- **Aggregation (in public preview):** Snowflake's aggregation policies give the data owner control over what can be done with their data even after it is shared with a consumer. Specifically, the owner can require a consumer of a table to aggregate the data rather than retrieve individual records. When creating an aggregation policy, the provider's policy administrator specifies a minimum group size; for example, the number of rows that must be aggregated together into a group. The larger the minimum group size, the less likely it is that a consumer could learn sensitive information from one query.
- **Projection (in public preview):** Snowflake's projection constraints control what data goes into the clean room and how data in the clean room can be joined with other data in the environment. Data owners can grant users access to a raw data table, but limit what columns can be included in the final result. For example, projection policies can be used to constrain identified columns such as names and phone numbers. The data consumer can still match records based on a particular value without being able to view that value. Projection constraints may be used within a data clean room or independently.

In addition to these controls, Snowflake Horizon features Differential Privacy, which has a unique ability to help you protect sensitive data while retaining analytical value.

Differential Privacy

Differential Privacy (currently in private preview) allows you to share data sets that include sensitive information like PII with a level of statistical confidence that the data does not reveal individual rows. Snowflake's Differential Privacy works by adding a controlled amount of randomness or statistical "noise" into queries on protected data. The added noise helps ensure that the results remain useful for meaningful analysis, but prevents (to a statistical degree of confidence) anyone from extracting row-level information from the underlying data. It protects the data from accidental leaks as well as targeted privacy attacks.

At the same time, Differential Privacy policies don't limit the value of the data by masking or removing individual identifiers. Data analysts can still query the data — including sensitive fields — at a granular level to understand trends and behavior. However, they won't be able to see row-level data or reverse-engineer sensitive information.


Sensitive data: Full granularity, unmasked, not anonymized
Differential Privacy Policy
✓ Analysts can see:
• Statistics and aggregates
• Feature engineering for AI/ML
✗ Analysts can't:
• See row-level data
• Identify sensitive entities
• Reverse-engineer proprietary info
Analysts
External / Internal

SNOWFLAKE DIFFERENTIAL PRIVACY ALLOWS YOU TO SHARE DATA SETS WITH SENSITIVE INFORMATION BY ADDING STATISTICAL "NOISE" THAT KEEPS THE DATA USABLE FOR ANALYSIS BUT PROTECTS SENSITIVE DATA FROM EXTRACTION.

Ultimately, you can use Snowflake's Differential Privacy to leverage sensitive data in your organization that was previously unavailable to analyze and share due to privacy concerns and regulations. This is especially relevant in the healthcare and life sciences as well as financial services sectors. You can apply interactive analytics and/or share the data internally or externally to drive strategic decision-making, improve business processes, and innovate new products or services — all while protecting your customers and sensitive information, and complying with privacy regulations.

ROKU MAKES DATA PRIVACY A PRIORITY WITH DATA CLEAN ROOMS

Privacy is a consumer-first concept at Roku, an American company that manufactures digital media players for video streaming, licenses its proprietary streaming software, sells channel subscriptions, serves targeted ads to viewers on its platform, and operates an ad-supported streaming channel. The company's primary consumer privacy goal is to provide consumers with transparency around its data collection and management.

Roku strives to build privacy controls that are clear, transparent and easy for customers to use, including a "privacy hub" where they can control how their data is used. Snowflake Native Data Clean Rooms allow collaboration with partners without any data leaving Roku's ecosystem, ensuring strong privacy protection for streamers.

Advertisers are a primary beneficiary of Roku's data clean room because it allows enrichment, analytics, measurement and more in a protected way. Roku can give advertisers access to current insights about Roku's first-party data insights, and the advertisers can choose and configure segments of their own data to bring into the clean room. The direct connection to Roku's platforms means advertisers can activate audiences in hours instead of days.

Watch this video to learn more about Roku's data clean room project.

INTEROPERABILITY

Governing content as it moves between clouds, platforms and regions can be tricky, especially because you must adhere to the unique rules of each cloud provider. Integrating across a customer's entire data estate to maintain a unified view of governance demands significant time and resources. With the rise in the use of open table formats such as Apache Iceberg tables, customers also don't want to be locked in to any one engine or platform. These challenges create complexity and can cause disruptions to the business and customers, potentially leading to financial consequences, operational interruptions, and security and compliance vulnerabilities.

Snowflake Horizon addresses these challenges with a unique governance approach that values and prioritizes interoperability. Snowflake Horizon's capabilities include:

- **A cross-cloud technology layer** that interconnects business ecosystems across regions and clouds
- **One-time definition of governance policies** and security measures that are universally applied
- Ability to **lift and shift governed data** from one environment to another with minimal effort and disruption
- **Pre-built integrations** by leading data catalog, governance and security partners to manage entire data estates both inside and outside of Snowflake
- **Integrations with Apache Iceberg**-compatible catlogs and engines to retain or gain flexibility

Snowflake Horizon empowers companies to reap these advantages with a robust suite of governance capabilities that offer simplified cross-cloud migration, pre-built partner integrations and enhanced flexibility with Apache Iceberg tables.

Work on any cloud to accommodate business needs

There are many reasons an organization might migrate to another cloud, including business consolidations, cost savings and access to certain technologies and tools uniquely available on a specific cloud provider. But a migration between cloud platforms can be laborious, disruptive and risky. Snowflake Horizon has the unique ability to empower consistent governance, including capabilities for cross-cloud business continuity.

Snowflake's cross-cloud technology layer, Snowgrid, interconnects business ecosystems across regions and clouds, enabling organizations to easily replicate databases, shares, governance policies, and even entire accounts across regions and clouds. Governance teams can apply any or all of the security controls — such as data access policies or user-defined tags — required to classify and protect sensitive data, and have that consistently migrated to another cloud environment.

CONSISTENT GOVERNANCE

With Snowflake, governors and stewards only need to define governance policies and security measures once, and these can then be applied seamlessly and universally — regardless of whether assets stay in one place or move to different regions, clouds or platforms. Consistent governance reduces the effort and risks associated with setting up and tracking these policies, which is especially important for highly sensitive data.

CROSS-CLOUD BUSINESS CONTINUITY

An organization might be planning a migration between Amazon Web Services (AWS), Google Cloud Platform (GCP) and Microsoft Azure. A consolidation of cloud vendors or platforms is not uncommon during a business merger or acquisition. Sometimes, technology leaders want the option to change their cloud vendor in the future. Whatever the reason, Snowflake makes migrating data in a governed way easy, allowing organizations to lift and shift from one environment to another with minimal effort and disruption. Snowflake Horizon's interoperability facilitates business continuity without impacting consumers and applications.

Easily integrate with catalog, governance and security partners to manage assets outside Snowflake

For organizations looking to empower data-driven decision-making, it's crucial to centrally manage your entire data estate. This involves not only optimizing existing investments, but also having the option to leverage metadata catalog, lineage, security and quality capabilities from leading solution vendors. To help organizations with these goals, Snowflake Horizon offers pre-built integrations with leading catalog, governance and security partners, enabling flexibility, efficiency and "better together" capabilities — all of which help customers better manage and govern their assets outside of Snowflake.

In addition, even though Snowflake offers best-in-class capabilities, customers may prefer the flexibility to use third-party governance solutions as they see fit. Snowflake works with these third parties to optimize integrations on core capabilities.

For more information on pre-built integrations by Snowflake Horizon's partners, visit the Snowflake Horizon Partner Ecosystem.

Connect to other Apache Iceberg-compatible catalogs and engines to gain flexibility

Snowflake Horizon allows organizations to connect their ecosystem both within and outside of Snowflake through integrations with other Apache Iceberg-compatible catalogs and engines. Customers can configure Iceberg tables to use either Snowflake or an external service such as AWS Glue as the catalog to track metadata.

To give customers the flexibility to organize their architecture based on their specific needs and preferences, Snowflake is rolling out three new features:

- **Catalog integration for Iceberg (in public preview):** Allows Snowflake to access, read and use metadata from AWS Glue and object stores.
- **Snowflake Iceberg Catalog SDK:** Allows Spark clients to read Snowflake-managed Iceberg tables. This new feature is a contribution by Snowflake to Apache Iceberg and is freely available for anyone to use and improve.
- **Iceberg Catalog REST API (in development):** Intends to allow any engine that follows the API's rules to read from Snowflake-managed Iceberg tables. Also, Snowflake will be able to read any other catalog that supports the Iceberg REST API.

CATALOG INTEGRATION FOR ICEBERG

Snowflake Horizon enables the democratization of data typically stored in various data silos through integrations with other Apache Iceberg-compatible catalogs, such as AWS Glue or directly from an object store. This expands the customer's access to data stored internally within Snowflake as well as data managed in various data lakes.

ICEBERG CATALOG SDK AND REST API

Modern architectures may consist of numerous compute engines that vary in specialization based on workload demands. Between structured to unstructured data processing and near real-time processing tools, an increasing number of compute engines work interchangeably over open table formats, eliminating the need to copy or duplicate data in proprietary formats while also providing a transaction-safe system through catalogs.

Snowflake Horizon provides a Java Catalog SDK that enables other processing engines to interact with all Snowflake-managed Iceberg tables. To expand the interoperability of processing engines, Snowflake intends to support the Iceberg Catalog REST API based on the open source specifications of the Apache Iceberg project. This would allow any engine that supports the Apache Iceberg REST API to interoperate with Snowflake-managed Iceberg tables.

Snowflake is also looking to roll out a new catalog integration to support the open source specifications of the Iceberg REST Catalog API. This new catalog integration would increase the flexibility to support other catalogs such as the most prevalent catalog, the Hive Metastore, as well as custom catalog implementations. Customers would be able to query and access their entire data estate from a single data management plane. Snowflake is committed to providing the flexibility and interoperability with open standards over the broad ecosystem of technologies in data estates.

.

HD SUPPLY SAVES 30% IN DATA TRANSFER COSTS

HD Supply, a leading wholesale distribution company, was acquired by The Home Depot in 2020. In the early stages of the acquisition, associates shared customer and supply chain data between the two companies' Snowflake accounts in Microsoft Azure. But when the combined business decided to move to GCP as its primary cloud platform, consolidating and migrating both accounts without disrupting the business became a priority.

HD Supply used account replication and failover features powered by Snowgrid, Snowflake's cross-region, cross-cloud technology layer, to complete the consolidation and cross-cloud migration in less than 60 days — achieving 30% savings in data transfer costs in the process.

To migrate from Azure to GCP, the company needed to migrate 20 TB of data across 25 databases on different clouds, ideally with little to no impact to data consumers. The final failover group setup and initial replication was performed in a single day. And because Snowflake features have the same functionality across clouds, the impact to data consumers was negligible, and users were able to access data easily and consistently throughout the process.

Read more about HD Supply's Snowflake story.

ACCESS

Access is the keystone of a modern cloud data platform. Accelerating access to content leads to faster iteration, shorter project development cycles, deeper collaboration, and better outcomes. But to reap these benefits, employees and partners must be able to find and share accurate, relevant content — and easily take action on it. With data estates constantly changing and growing, organizations face an ongoing battle to understand what's in all their content, if it is appropriately governed and relevant to their business initiatives, and grant access to it.

Snowflake Horizon enables companies to achieve these goals by offering a robust suite of governance capabilities for customers to classify, share, discover and immediately act on content across regions and clouds, including:

- **Tagging and auto-classification features** that use automation to remove the "button pushing" burden from data stewards, and make it easy to get a comprehensive view of the state of tagging across your data environment
- **AI and LLM-powered search capabilities** to help you find and query content across a broad variety of sources by asking questions in simple English
- **Direct data sharing and cross-cloud auto-fulfillment** across internal business units and with external partners

These features work together with the other elements of Snowflake Horizon to provide end-to-end governance coverage that helps organizations efficiently protect and manage content access, while still ensuring that content is available for sharing, analytics and monetization.

Identify, tag and monitor sensitive data

Tagging and classification are initial steps toward discovering and understanding your data. Today's tagging efforts, however, must go beyond just adding a sensitive data tag to PII. Adding detailed metadata, such as documentation links, a semantic description of the data, or even information about the geographic location of customer data is a significant factor in how well your search and discovery processes perform.

Metadata-driven tags allow organizations to classify objects as well as track PII, monitor usage, and apply masking and data protection policies. This is especially important as organizations embrace LLMs to expand the capabilities of AI models. LLMs thrive on semi-structured and unstructured data, so to prevent unintentional exposure of PII, organizations must understand what information is in the documents, text fields, records and other data sources being used to train the LLM. With this knowledge, data stewards can quickly and effectively govern what resources are appropriate for LLM training, which projects can access the LLM, and who can access the results of the models (which may also include PII).

Snowflake Horizon brings modern, machine-augmented tagging and classification capabilities to today's increasingly large data environments. Customers can now better understand what's actually in their data — and reduce the risk of a reputation-damaging data leak or financially devastating regulatory fine.

OBJECT TAGGING AND ENHANCED CLASSIFICATION

Snowflake Horizion's data tagging and classification features help you define what sensitive data means to your organization and identify where PII exists in your data environment, if it's appropriately tagged, and where it came from.

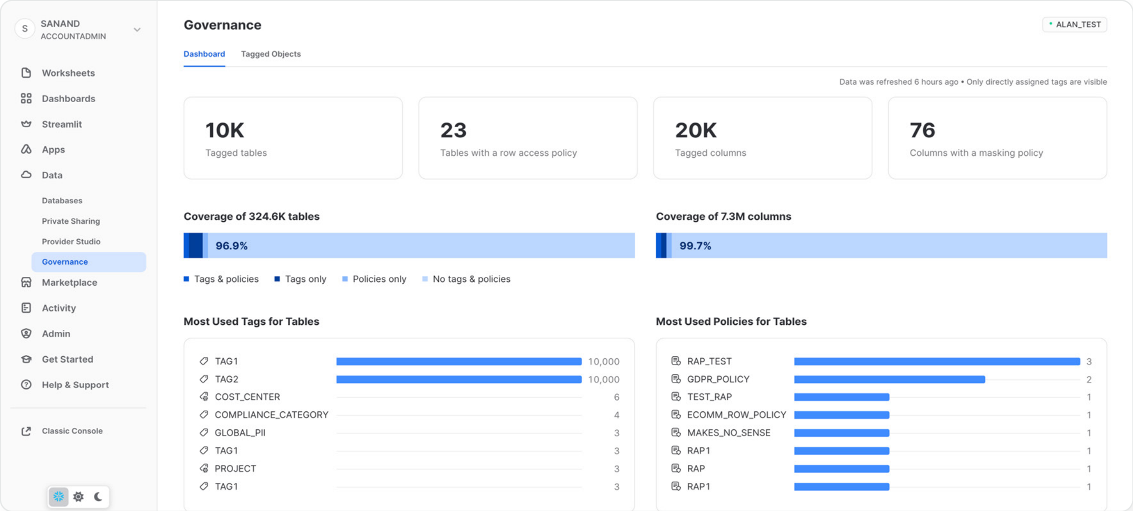
- **Object tagging** involves assigning metadata to Snowflake objects to describe the type of data stored in a table or column. Once defined, you can apply the tag to appropriate objects based on your governance strategy. Tags stay with data when it is replicated, reducing the chance of PII being exposed when data is copied or moved.
- **Sensitive data custom classification capabilities (in private preview)** allow you to create your own classifiers to identify and tag data that your organization specifically considers to be sensitive. You can write your own pattern matching on both column names and the data records in those columns. For example, a manufacturing organization could write pattern matching to identify part numbers, while a healthcare provider could identify patient IDs.
- **Data Classification UI (in public preview)** lets you start a data classification job for an entire schema or a subset of tables within it through an intuitive UI. Choose when you want to review and apply the classification results. You can also run classification with auto-tagging, which allows you to "auto-apply" high-confidence Snowflake classifiers to objects.

By adding automation to tagging and classification processes, Snowflake helps reduce the amount of manual tagging work data governors and stewards must do, and makes it easier for them to determine if a particular data product should be shared.

EFFICIENTLY ADDRESS TAGGING AND ACCESS IN A SINGLE UI

The Snowflake Data Governance Interface brings it all together. Designed to be used by business users with less SQL exposure as well as data stewards with many governance tasks to complete, this Snowsight dashboard delivers quick, proactive object identification so users can take immediate action. Data stewards and data governors can:

- View the state of sensitive data in their account, zeroing in on a specific database or schema to get additional details
- Monitor their schemas from a single location and generate detailed reports for audits with a few clicks
- Get a comprehensive, detailed list of tagged and untagged objects across databases and schemas
- Drill into untagged objects and take immediate action from within the UI to apply tags or assign masking and access policies



Governance

Data was refreshed 6 hours ago • Only directly assigned tags are visible

10K
Tagged tables

23
Tables with a row access policy

20K
Tagged columns

76
Columns with a masking policy

Coverage of 324.6K tables
96.9%

Coverage of 7.3M columns
99.7%

■ Tags & policies ■ Tags only ■ Policies only ■ No tags & policies

Most Used Tags for Tables

TAG1	10,000
TAG2	10,000
COST_CENTER	6
COMPLIANCE_CATEGORY	4
GLOBAL_PII	3
TAG1	3
PROJECT	3
TAG1	3

Most Used Policies for Tables

RAP_TEST	3
GDPR_POLICY	2
TEST_RAP	1
ECOMM_ROW_POLICY	1
MAKES_NO_SENSE	1
RAP1	1
RAP	1
RAP1	1

THE SNOWFLAKE DATA GOVERNANCE INTERFACE PROVIDES AN AT-A-GLANCE, EASY-TO-UNDERSTAND VIEW OF THE STATE OF SENSITIVE DATA ACROSS AN ORGANIZATION.

Discover, query and easily interact with relevant content

When data analysts need to answer a question, finding the right content is half the battle. While internet searches deploy highly evolved indexing to find appropriate answers even when your search term is misspelled or vague, the same hasn't necessarily been true when searching data repositories. If you ask for "sales data" and the exact-string-match search functionality doesn't bring up a table named "sales_data," you're missing out on potentially crucial information.

In Snowflake Horizon, search is not a commodity feature — it's transformative. Snowflake's Universal Search (currently in public preview) is an LLM-powered search that adds a new level of intelligence to the search experience by returning results across sources that users may — or may not — be familiar with, including:

- Database, schema, and table objects and views in your Snowflake account
- Data products such as data and apps on Snowflake Marketplace
- Snowflake documentation pages
- Snowflake community Knowledge Base articles

Built on search technology acquired through Neeva, Universal Search helps streamline the process of discovering and accessing content by delivering:

- **Natural language queries:** Use conversational language like "sales opportunities that came from partner referrals" to focus your search request, or apply classic keyword search terms like "sales opportunities."
- **Metadata-based indexing:** Universal Search indexes object metadata — the object name, table name, column names, tags, comments and descriptions on the table or individual columns — and uses elements like the popularity of tables to score relevance.
- **Ranked results for fast action:** For each search query, Universal Search will understand the natural language semantics, correct spelling errors, perform synonym expansion (a search for "billing" might be expanded to include "invoices"), and do ML-based scoring to return a ranked list of the most relevant results so you can quickly decide which action to take.

Snowflake Copilot (currently in private preview) presents another way to interrogate the Universal Search index: conversationally. Built on Universal Search, analysts can ask Copilot a question in natural language and it will answer by providing the most relevant database objects or generating a SQL query. Refining the query is as simple as continuing the conversation and asking follow-up questions. This saves time for analysts and allows them to easily access a broad and deep set of relevant data.

Share direct access to content across teams, business units and partners

Once you've identified, tagged and classified your content, you have a better understanding of what you have, where it exists, and who can (and cannot) access it depending on how and when the content is being used. To reap the full value of this data, it needs to be shared so it can be used to drive and inform business decisions. But how do you do this without complex ETL or FTP integrations and without risking exposure of the sensitive data you've so carefully classified?

Snowflake Horizon's end-to-end governance capabilities allow you to securely share access to live data and Snowflake Native Apps (generally available on AWS and Azure, in development on GCP) within and across your organization and with business partners and customers. Listings help you significantly streamline internal operational processes by removing the need to copy and move data or deal with ETL while maintaining governance over shared data.

Instead of just locking down your data, Snowflake opens up the potential for internal and external collaboration — even around sensitive data — while helping to keep data secure and protected. It facilitates deeper collaboration between teams and partners using a variety of Snowflake innovations, including:

- **Listing discovery controls** allow you to share data directly with particular business units and partners by specifying who can discover a listing and the associated content. Rather than setting up direct shares or private exchanges that often require replicating and sending data between regions, listings give both data providers and consumers a simplified, richer collaboration experience. Listings deliver insights into who is accessing your data and how they're using it through robust programmatic and visualized analytics, further strengthening the governance of the shared data.
- **Cross-cloud auto-fulfillment** makes up-to-date data available by automating replication and fulfillment across clouds and regions. All instances of data or apps are updated consistently based on your preferred syncing schedule. No more concerns about teams having different versions of a data set or partners working with an out-of-date app.
- **Monetization** opens opportunities to customize your data products and pricing by defining the users and companies with whom you want to share data and understanding how they interact with your listings. With multiple pricing models available, including usage-based pricing, you can use Snowflake Marketplace to turn your enterprise data assets into a new revenue stream.

MARKETSCAN EXPEDITES TIME TO VALUE FOR CLIENTS THROUGH NEAR-INSTANT ACCESS TO REAL-WORLD DATA

MarketScan by Merative provides real-world data to life science companies and other organizations to support research and regulatory submissions. Before Snowflake, MarketScan manually shared data via Amazon S3 or FTP, which required considerable planning and coordination. Some clients lacked in-house data engineering expertise, delaying time to first insight by up to three months.

To improve data collaboration, Merative sought out a more SaaS-like delivery model for MarketScan data. The company now uses Snowflake Marketplace as an innovative way to allow clients to securely connect to de-identified, longitudinal, patient-level and specialty data. With Snowflake, Merative's clients access the latest MarketScan data 10x faster than before and can save up to 60% of their ingestion and processing costs.

Using Snowflake for data collaboration helps MarketScan deliver data faster and frees up technical staff to focus on higher value tasks. With Snowflake Marketplace as their primary data sharing mechanism, MarketScan can take advantage of the automation and access benefits to drive additional efficiency for clients — and capitalize on an opportunity to increase demand in life science organizations that also use Snowflake.

Read more about **MarketScan's data sharing success story**.

CONCLUSION

With Snowflake Horizon, organizations can leverage a unified set of compliance, security, privacy, interoperability and access capabilities to build a trusted, secure data foundation. To learn more about how to ensure strong governance in your organization, visit **snowflake.com/en/data-cloud/horizon/**

ABOUT SNOWFLAKE

Snowflake enables every organization to mobilize their data with Snowflake's Data Cloud. Customers use the Data Cloud to unite siloed data, discover and securely share data, power data applications, and execute diverse AI/ML and analytic workloads. Wherever data or users live, Snowflake delivers a single data experience that spans multiple clouds and geographies. Thousands of customers across many industries, including 691 of the 2023 Forbes Global 2000 (G2K) as of January 31, 2024, use Snowflake Data Cloud to power their businesses. Learn more at **snowflake.com**

Forward-Looking Statements: This document contains express and implied forward-looking statements, including statements regarding (i) Snowflake's business strategy, (ii) Snowflake's products, services, and technology offerings, including those that are under development or not generally available, (iii) market growth, trends, and competitive considerations, and (iv) the integration, interoperability, and availability of Snowflake's products with and on third-party platforms. These forward-looking statements are subject to a number of risks, uncertainties, and assumptions, including those described under the heading "Risk Factors" and elsewhere in the Quarterly Reports on Form 10-Q and Annual Reports of Form 10-K that Snowflake files with the Securities and Exchange Commission. In light of these risks, uncertainties, and assumptions, actual results could differ materially and adversely from those anticipated or implied in the forward-looking statements. As a result, you should not rely on any forward-looking statements as predictions of future events.