



# DATA TRENDS 2023

Four essential trends redefining the way modern companies  
succeed with AI, automation, and more

# DATA AND AI ARE REINVENTING THE WAY WE DO EVERYTHING.

The world is changing fast, and one of the most powerful drivers of change is data. Data is reshaping research, policy, innovation, and more. Generative AI is prompting every industry to rethink core workflows and processes—catalyzing a fundamental shift across software and enterprises. Data helps cybersecurity teams respond at faster-than-human speeds to an ever-rising number of threats that are also using fast and sophisticated data-driven technologies. Data can be used to improve everything from supply chain logistics to the proper stocking of retail shelves to accelerated development of medicines and vaccines. And that's today; whole workflows, categories of software, and jobs will be transformed in the coming years.

Because the Snowflake Data Cloud is fundamental to how thousands of companies are able to get those insights, we were able to look at how they're optimizing for the future. This wasn't research into how business or technical leaders *feel* about their data operations, nor about what they intend to do or believe is working—we looked at actual usage data throughout the past year to see trends in how organizations leverage their data, from governance issues to the programming languages they're using.



## METHODOLOGY

The data in this report covers the 12-month period ending Jan. 31, 2023, referred to as “this year” or “the year,” to align with Snowflake’s 2023 fiscal year. We examined data usage of roughly 7,800 Snowflake customers, some of them longtime Snowflake users, and others only recently having joined the Data Cloud. Note that Snowflake’s customer base grew 31% in the 2023 fiscal year, which provides a baseline of comparison for statistics identifying trends that outpaced this overall growth.

**HERE ARE FOUR  
KEY TRENDS  
THAT SHOW HOW  
ORGANIZATIONS ARE  
USING THEIR DATA  
AND EMBRACING AI.  
IN OTHER WORDS,  
THESE ARE  
TRENDS WITH REAL  
MOMENTUM.**



**TREND 1:**

Companies are connecting data and AI everywhere they can.



**TREND 2:**

State-of-the-art companies are bringing their work (and AI) to the data, not vice versa.



**TREND 3:**

Governance is even more important in the new AI era.



**TREND 4:**

Companies are embracing automation and expect a fully managed platform.



## TREND 1:

# COMPANIES ARE CONNECTING DATA AND AI ACROSS THEIR BUSINESS ECOSYSTEMS, GLOBALLY AND ACROSS CLOUDS.

The importance of connecting data is often talked about as a key IT imperative. In [Data and Analytics Trends 2023](#), the Info-Tech Research Group identifies “democratizing real-time data” as strategically crucial. The report also calls out the importance of data marketplaces that share data for collaborative purposes, and notes Amazon, Microsoft, and Snowflake as leaders in this trend.

When [McKinsey identified seven characteristics](#) that will define the data-driven enterprise of 2025, No. 1 was “data embedded in every decision, interaction, and process.” The next two were “data is processed and delivered in real time” and “flexible data stores enable integrated, ready-to-use data.” All of this points to environments without data silos or the lengthy resource delays of collecting and cleansing the data set. And the imperative, across every industry, to use that data to power AI models makes these characteristics even more vital.



**We looked at how our users are interacting with the Snowflake Data Cloud to see whether real-world behaviors match predicted trends. The answer, overwhelmingly, was yes.**

### **Cross-cloud growth**

Fewer companies are confining their data to one cloud or region; many are using a cross-cloud strategy for business continuity, resilience, and collaboration. Over the year, the number of Snowflake customers operating across the three leading public cloud providers (Amazon Web Services, Microsoft Azure, and Google Cloud) grew 207%. This exacerbates a growing need for applications to share data globally and across clouds.

### **Resilience**

Organizations are increasingly emphasizing business continuity strategies to mitigate risk. Cross-cloud replication can facilitate seamless failover from one cloud to another, without disrupting the business. On Snowflake, the number of Forbes Global 2000 customers as of Jan. 31, 2023, doing cross-cloud replication increased 58% year over year.

### **Collaboration**

Data that lives in silos is not living up to its potential. Throughout the year, we saw increased collaboration on data across regions and across cloud providers.

Snowflake allows concurrent sharing of data where it resides, which dramatically increases efficiency by cutting out the need to extract, transform, and load that data before you can work with it. Billions of jobs were run against shared data during the fiscal year, with the number of jobs run growing 93% comparing Feb. 1, 2022, to Jan. 31, 2023.

A measure of collaboration that is unique to Snowflake is stable edges. In brief, an “edge” is a data share between a provider and consumer of data. A “stable edge” is one that, over two successive three-week periods, has produced at least 20 data transactions in each period. This indicates that the specific collaboration this data sharing facilitates has ongoing value to the organization. Stable edges increased 93% over the year, suggesting that Snowflake users nearly doubled the usage of shared data on an ongoing basis to improve outcomes.

**In short, organizations continue to uplevel how they connect their data and the data of others to improve processes and insights. We’re seeing not only experimentation, but ongoing adoption of these successful innovations.**

In 2023, the number of organizations with data across all three major public clouds grew by

**207%**

### **COMPLETE DATA DRIVES BETTER DECISIONS**

Medical technology maker Siemens Healthineers acquired and merged with Varian Medical Systems, a medical device company known for treating cancer. [Working with Snowflake](#), Siemens Healthineers made integrated CRM data available globally. Whereas this process previously would’ve taken weeks or months, it took only about 20 minutes for the data from a Snowflake Azure instance in the United States to be replicated and available in Snowflake Azure in Europe. This let the sales team analyze the data right away to derive new insights, drive global decision-making, and identify cross-selling opportunities.



## TREND 2:

# THE NEXT STEP OF “BREAKING DOWN DATA SILOS” IS TO BRING YOUR WORK, INCLUDING AI, TO ALL OF YOUR DATA. STATE-OF-THE-ART COMPANIES ARE DOING IT ALREADY.

Another source of silos is that data is generated in many formats, and lands in different specialized systems to be consumed by disparate downstream teams. Bringing together different formats and types (structured, unstructured, semi-structured) is a persistent challenge. Companies have been making fitful progress on these challenges for more than a decade. But storing all your data in one place is less meaningful if you have to pull out and prep discrete data sets for each kind of job you want to do. The next horizon—one that’s essential to the dawning era of generative AI—is being able to do meaningful work with all that data together.

Because Snowflake lets users analyze any of their data, in the same platform and with the same engine, we can see how organizations embrace the ability to bring their work to all their data. This work can include building pipelines to process data, training machine learning models, creating analytic queries, dashboarding, and even powering entire applications. As of January 2023, Snowflake has more than 800 companies registered in the Powered by Snowflake program, which helps them build, support, and

### PYTHON + SQL: A POPULAR PAIR

When users bring their work to their data, rather than vice versa, how are they doing it? While SQL remains the most popular language on Snowflake, the most popular programming language used with [Snowpark](#), our developer framework, is Python. No surprise, since Python’s versatility, ease of use, and large community make it popular with programmers everywhere.

- In February 2023, Python accounted for nearly 88% of all jobs run on Snowpark.
- In addition, the Python Connector was the most popular tool for running DML jobs on Snowflake over the year.

Our takeaway is that organizations need a platform that allows many languages, including SQL and Python, to be used by the same engine, with the same governance, against the same copy of data.



scale their applications in the Snowflake Data Cloud. One exciting new trend is how more and more of these SaaS providers are architecting their applications to connect to a customer's data platform instead of re-siloing data into their own managed data store. During the year, the number of connected applications in the Data Cloud grew 285%.

We're seeing that once companies have all their data in one place, they can—and want to—do more with it. Snowflake users collectively run billions of jobs per day on the data within the Snowflake Data Cloud. Over the year, we saw the number of jobs in the Data Cloud grow by 64%—more than double our 31% customer growth, indicating that even as we simply add users, those users are finding more and more ways to bring work to their data.

They're using SQL, Python, Java, Scala, and other languages to work with data within the Data Cloud, connecting to a rich repository of data to drive improvements and insights. In January of this year, one customer ran more than 900 million jobs in that month alone—averaging out to about 20,000 jobs a minute.

**Our conclusion is that once you make it easier for someone to work with all their data for a variety of workloads, they'll do so, unlocking new value from that data. Especially when it can be supported at enterprise scale, where concurrency is a requirement. We expect more and more companies to find ways to eliminate the movement of data and to bring more work to the data itself in the AI-driven years ahead.**

Over the year, the number of jobs run in the Data Cloud rose **64%**

## **ENERGIZING DATA SCIENCE**

EDF is a leading energy supplier in the UK, and is Britain's biggest generator of zero-carbon electricity. [EDF uses Snowflake](#) as a central repository for all customer data. Working with such data can be challenging. Prior to Snowflake, the company's data lake team had to provide extracts for data scientists to work with, which was slow-moving and came with a lot of complexity. Data scientists had to then manage the security and governance of that data themselves.

Rebecca Vickery, EDF's data science lead, says that Snowpark, with its support of Python and SQL, lets business users manipulate that data where it lies, deploying end-to-end machine learning to uncover the insights that make customers' lives easier.

**"The benefits of being able to run data science tasks, such as feature engineering, directly where the data sits, is massive," she says. "It's made our work a lot more efficient and a lot more enjoyable."**



## TREND 3:

# GOVERNANCE IS EVEN MORE IMPORTANT IN THE NEW AI ERA.

Data governance—how an organization understands and protects its data, and puts it to use—defines the roles, processes, and policies for interacting with data. Effective governance helps you develop business insights from trustworthy data and helps ensure regulatory compliance.

The compliance part is a big deal. In the past several years, regulation and standards around data protection have risen sharply. The European Union's General Data Privacy Regulation (GDPR) and the California Consumer Privacy Act (CCPA) get an outsized share of attention, but there are many others. Canada has PIPEDA (the Personal Information Protection and Electronic Documents Act) and Mexico has the Ley de Protección de Datos Personales. In Brazil it's LGPD (Lei Geral de Proteção de Dados Pessoais); in Singapore it's PDPA (the Personal Data Protection Act), and South Africa has POPIA (the Protection of Personal Information Act).

And that's just off the top of our heads. Except for Canada's, all those regulations have come into effect in the past decade, and each regime brings the threat of significant fines for non-compliance.

But effective governance is about more than checking regulatory boxes. Strong data governance is an enabler, even an accelerator, for getting full value from data. As modern enterprises use generative AI and LLMs to extract insights from their own data, a single, consistent governance model is indispensable. For both reliability and security, the full breadth of an organization's data must be efficiently, properly governed.

A core aspect of good data governance is to ensure that only people in the right, necessary roles are able to access a given data set. An added complexity is that different roles might need access to the same data set, but with differing levels of visibility. Organizations need automated and dynamic controls and policies, including classification, tagging, masking, and granular role-based access controls, to enforce this at scale.





Any trendwatcher or prognosticator will tell you that organizations are increasingly concerned with meeting data governance requirements. (That Info-Tech report notes “adaptive data governance” as a key trend, while a McKinsey report [advises companies](#) to establish an enterprise-wide data governance program to help themselves thrive in economic uncertainty.) But we wondered what organizations are actually doing.

So we took a look at our own user base. Snowflake offers a wide range of native governance policies and controls that can be consistently enforced across all data, regardless of region and cloud. As an example, users can employ dynamic masking policies that conceal sensitive data down to fine-grained levels from roles that aren't permitted to access that data. We found that our users are increasingly applying these controls to understand and protect data at scale. Across the Snowflake Data Cloud, the number of applied dynamic masking policies active for capacity customers had grown more than 205%, comparing Jan. 31, 2023, to Feb. 1, 2022. Again, that's six times the growth of our customer base, suggesting that users globally are becoming more rigorous in their application of such governance policies.

Furthermore, after releasing the feature in 2021, we saw meteoric growth in object tagging, which helps add context to data to better use and protect it, as well as trigger automations and actions. Users may use tags to track personally identifiable information in a data set, and then understand who has been accessing those sensitive data sets. The number of such tags applied by Snowflake customers increased nearly 40x comparing Jan. 31, 2023, to Feb. 1, 2022.

**Our takeaway is that as privacy becomes more important, organizations will turn away from siloed environments with separate governance controls, favoring a single platform that makes it easy to know and protect your data and your AI models.**

### **SQUARE: SMARTER, MORE EFFECTIVE DATA GOVERNANCE**

Financial services platform [Square](#) [uses Snowflake](#) to enable secure, governed access to data. Square's data platform team relies on Snowflake to answer complex questions about data lineage, data accessibility, and data versioning. Anomaly detection algorithms monitor Square's data lifecycle to avoid bad data and ensure data quality. The company is progressing toward self-service data governance, in which data customers can take care of their own data rather than waiting on a data governance team.

# 40x

The number of tags applied by Snowflake customers increased nearly **40x** comparing Jan. 31, 2023, to Feb. 1, 2022.



## TREND 4:

# COMPANIES ARE EMBRACING AUTOMATION AND EXPECT A FULLY MANAGED PLATFORM.

Increasingly, customers are taking advantage of automation capabilities, enabling greater efficiency and minimizing operating costs. This will be increasingly important: Running massive AI models requires scale, expertise and planning for an unpredictable amount of resources. Human speed isn't fast enough; every organization must look to take its approach to automation to the next level.

A basic example is resizing compute resources to right-size the resources and improve query performance. In January 2023, Snowflake performed millions of automated warehouse resize events a day to adapt to customer needs. We saw a 71% rise in such activity comparing Jan. 31, 2023, to Feb. 1, 2022.

The shift to greater use of automation creates a new challenge for CIOs, CDOs, and data platform administrators: how to establish financial governance. Data platform owners want to track use of resources by department or purpose, monitoring consumption to help align costs to value.

The cloud has unlocked abundant resources, allowing teams to deliver value faster. Compute can be spun up and spun down to meet the organization's changing needs rather than leaving IT teams trapped in a slow-moving world of fixed-capacity resources. In this new world, the mission for data platform owners is to establish their internal practices around cost optimization and financial governance to help executives connect cost to value.

Automation will not only power AI, it will be powered by it. An [EnterprisesProject article about automation trends](#) called out cloud resource optimization, edge computing, and DevSecOps as areas that would benefit from AI-driven automation in the coming years.

**Judging from what we see in the Data Cloud, data and IT experts are more than ready to take advantage of these automations to maximize performance and minimize costs.**

### VERADIGM: AUTOMATION DRIVES SCALE AND INNOVATION

Veradigm, a technology company that delivers care and financial solutions to healthcare providers, [turned to Snowflake to modernize](#) its data environment. Snowflake's multi-cluster shared data architecture, with automatic scaling of storage and compute resources, eliminated Veradigm's performance issues while lowering costs. The fully managed infrastructure and near-zero maintenance helped Veradigm's data team support additional data use cases without increasing head count.

# ELEMENTS OF DATA AND AI SUCCESS

Each of these four trends presents its own opportunities to form strategies that address business needs and the competitive landscape. Below are essential recommendations derived from both the individual trends and a more holistic consideration of how leading organizations are embracing new possibilities to use their data to improve decision-making and accelerate positive outcomes.



## REMOVE THE BARRIERS TO COLLABORATION

Many organizations have spent years trying to eliminate silos by bringing data together in one place. But their processes or tools—the actual work they do with the data—create new silos that inhibit collaboration and reduce efficiency. Whether you're collaborating with global teams, different business units, or with third-party providers, you should be able to work as seamlessly with your data as though you're all in the same room, on the same systems. Modernize your data environment with an eye toward making data secure and shareable, and toward reducing wait time. Teams should be able to access live data in the moment, to drive more powerful insights between analyst teams, data scientists, and others.



## LET TEAMS WORK IN THEIR PREFERRED LANGUAGE

Technical people almost always have preferred languages, and they'll gravitate to the tools that are designed for Python, SQL, Scala, etc. This can create translation problems, but the answer is not to

force your people to stop using the languages and tools they love. (They probably won't, and it might hurt retention of hard-to-replace talent.) Instead modernize your environment to allow people to work with their preferred tools and languages. You need to give them that flexibility while also providing shared governance, proper performance capabilities, etc. Otherwise, you're inevitably re-siloing your data through incompatible, inflexible work processes.



## TURN GOVERNANCE INTO AN AI ACCELERATOR

The purpose of governance is to keep data secure and manage access in order to comply with relevant policies and regulations. The easiest way to do that is to lock all data down and minimize access to it, which can also minimize the value of the data. Refocus your data governance regime around safely enabling the use of data. Don't create multiple copies of the same data set, with conflicting governance policies. Ensure that everyone who needs to can access the same canonical data set, but with role-based dynamic masking that makes sure each user sees only what they should be

permitted to see. Use automation to simplify governance of data across systems and clouds so that data is managed from one place rather than ad hoc across your entire ecosystem. And make sure that governance is considered up front, as data is ingested and uses are planned, rather than trying to securely manage it as an afterthought. All this ensures that the data is protected, while allowing the value to be realized.



## MINIMIZE TOTAL COST OF OWNERSHIP OF YOUR DATA SYSTEMS THROUGH AUTOMATION

As organizations scale in terms of data volume and overall complexity, it quickly becomes impossible to optimize your data environment through manual processes. Fortunately, machines scale when humans can't. Automating efficiencies in terms of how resources are managed not only prevents costly human error (like the dev environment no one remembered to spin down), but it also frees your teams to build rather than sink time into managing resources, installing upgrades, and performing other maintenance. Efficiency through automation prevents waste of resources and ensures more value through greater productivity.

# NEXT STEPS

Learn more about how Snowflake can help you improve governance and collaboration through automation and AI—and get more value from your data.

## OPTIMIZE COST AND PERFORMANCE

[Learn more about](#) how the Snowflake Data Cloud helps companies optimize performance and minimize costs.

## FIVE STEPS TO SUCCESSFUL DATA GOVERNANCE

For more guidance on Trend 3 in this report, [read this ebook](#) to discover how to define a data governance strategy and scale it as your data grows.

## OPERATE AT GLOBAL SCALE WITH SNOWGRID

[See how](#) Snowflake's unique cross-cloud technology layer, Snowgrid, helps global organizations overcome challenges in terms of collaboration, governance, and business continuity to maximize the value of their data.





# 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, and execute diverse 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 573 of the 2022 Forbes Global 2000 (G2K) as of January 31, 2023, use Snowflake Data Cloud to power their businesses.

Learn more at [snowflake.com](https://www.snowflake.com)



© 2023 Snowflake Inc. All rights reserved. Snowflake, the Snowflake logo, and all other Snowflake product, feature and service names mentioned herein are registered trademarks or trademarks of Snowflake Inc. in the United States and other countries. All other brand names or logos mentioned or used herein are for identification purposes only and may be the trademarks of their respective holder(s). Snowflake may not be associated with, or be sponsored or endorsed by, any such holder(s).