snowflake®

DATA TRENDS 2024

7 Ways Leading Organizations Are Building Toward Advanced AI Success

TABLE OF CONTENTS

EXECUTIVE SUMMARY

We looked at how more than 9,000 Snowflake accounts adopted features and capabilities of the Data Cloud over the previous fiscal year to reveal trends, both in terms of the foundational development of data infrastructure and those users' first moves into advanced AI. Generally, we compared January 2023 to January 2024 to align with Snowflake's fiscal year, except in cases where features went into public preview during the year. In those cases, we compared the first full month in public preview to January 2024. For the full methodology, **see the appendix.**

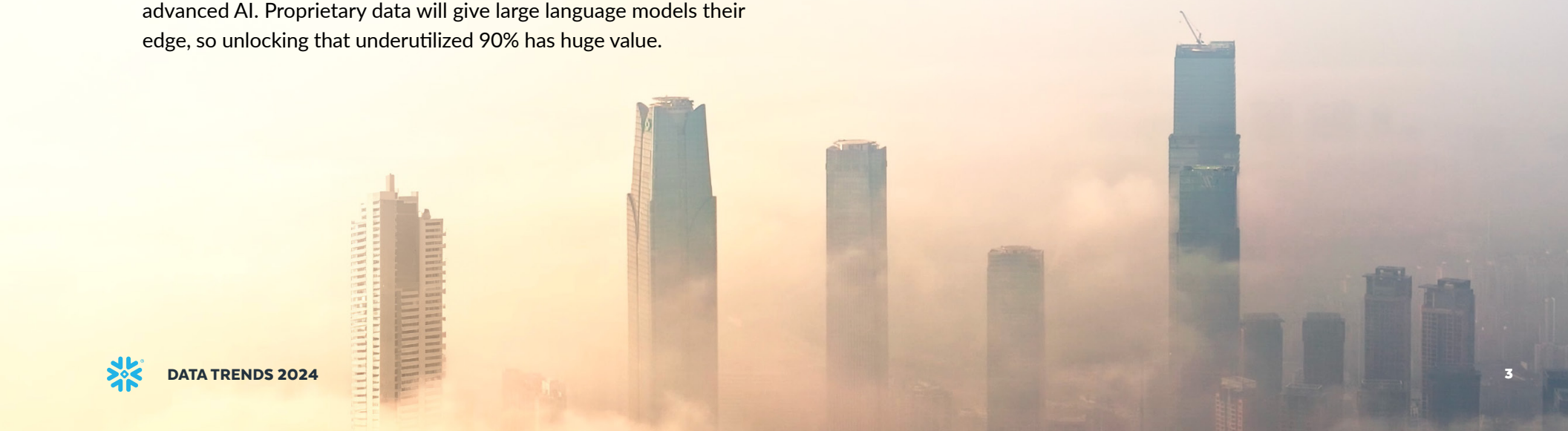
Highlights from this report include:

FIRMING UP THE DATA FOUNDATION

1 **Python is the language of choice for AI programming.** With its ease of use, active community and ecosystem of libraries and frameworks, **Python use grew 571%, considerably more than** any other language year over year. Python skills will be increasingly essential to development teams as they venture into advanced AI.

2 **Enterprises are finally tapping their unstructured data.** Most data—as much as 90% by some estimates—is unstructured videos, documents and more. We saw **processing of unstructured data grow by 123%.** That's good news for many uses, not the least of which is advanced AI. Proprietary data will give large language models their edge, so unlocking that underutilized 90% has huge value.

3 **Enterprises are getting fine-grained about their data governance.** We're seeing not just more governance measures applied to data; we're seeing a more refined approach as organizations embrace a wide range of tagging standards and features. The takeaway: While usage of every data governance feature rose 70%-100%, **the number of queries against protected objects is up 142%.** Governance is not about locking down your data, it's about making it more available for secure, authorized uses, and we're seeing exactly that.



MAKING AI ACCESSIBLE

1. **The democratization of AI is here.** A major promise of AI is that it will make technology available to less technical users. We've empowered that through the machine learning functions of Snowflake Cortex, and since public preview of key features began in June 2023, we've seen **the number of active accounts adopting ML-based functions grow by 67%.** That opens up more possibilities because data scientists and other experts are no longer a bottleneck.
2. **The LLM explosion is happening now—probably at your office.** What bottleneck? In the last fiscal year in the Streamlit developer community, we saw **20,076 unique developers work on 33,143 LLM-powered apps.** That means that the future filled with the power of AI is here. It may not be evenly distributed yet, but it's here.
3. **The chatbot is on the rise.** Single-text input LLM apps may be easier to make, but they don't allow refinement through natural conversation. For that you need chatbots, and increasingly that's what the devs are making. From May 2023 through January 2024 in the Streamlit community, **chatbots went from 18% of LLM apps to 46%.** And climbing.
4. **Enterprises want apps and data within a unified data platform.** We make it possible for users to build applications within our data platform, where their data resides, via the Snowflake Native App Framework. Maybe it's the ease of use or the single source of truth. Maybe it's the security and governance advantages. But the data shows that people want to bring the work to the data. **The number of Snowflake Native Apps grew 311%, and the use of those apps is up 96%,** based on January 2024 utilization compared to July 2023 (Snowflake Native Apps went into public preview on June 27, 2023).

ADVANCING IN THE AI AGE

We're now a year and a half into the generative AI era, and things are only accelerating. OpenAI's release of ChatGPT and then GPT-4, Meta's decision to open source Llama and Llama 2, and a host of other announcements and innovations around the application of advanced AI have stirred more excitement and driven real progress in the development and enterprise adoption of large language models.

Tremendous opportunities and challenges lie ahead, and as we analyzed use of the Snowflake Data Cloud to understand the latest trends around data and technology, our chief interest was around how enterprises are preparing for an unfolding era in which advanced AI accelerates and transforms how they do business.

The Snowflake Data Cloud encompasses data, models and applications from thousands of organizations across many industries. Looking at how they work within the platform, including which features they use, paints a vivid picture of the decisions being made to deal with current challenges and prepare for future success.

A lot of industry research surveys executives and practitioners, asking them to estimate things such as what percentage of their data is unstructured, or to describe how confident they feel about their approach to data governance.

This report didn't ask anyone's opinion. Instead, we looked at how enterprises worldwide are making decisions and applying their resources to leverage their data.

Through that lens, a picture emerges about how the modern, data-forward enterprise is shaping its data strategy on the cusp of an AI revolution. In short, business and technology leaders at these organizations are preparing for the future. They are taking initial steps into the world of large language models and generative AI. More importantly, they are fortifying their data foundation.

While the specific technologies around advanced AI—the algorithms and apps—are powerful, they don't work alone. To be successful, a business must build the shiny, new AI technology on top of a solid stack of organizational practices and technologies to ensure a company's data is available, secure and properly governed. In other words, the LLM is the dessert, while a solid data infrastructure is the main course.

In our predictions report for 2024, our in-house experts advised that the proper response to the new AI age is not to desperately create a new data strategy, but to accelerate the same solid, thoughtful practices you were following before you ever heard of ChatGPT.

When we look at how Snowflake users are working with their data, we see exactly that: a focus on silo-busting, refining governance practices, and finally coming to grips with the flood of unstructured data. For starters.

> **"The generative AI era does not call for a fundamental shift in data strategy. It calls for an acceleration of the trend toward breaking down silos and opening access to data sources wherever they might be in the organization."**

—JENNIFER BELISSENT

Principal Data Strategist, in **Snowflake Data + AI Predictions 2024**

CEMENTING THE DATA FOUNDATION

Organizations are doing a lot to make more data securely, appropriately available to today's tools and applications as well as tomorrow's (or next week's) AI advance. At the foundation layer, we've identified the following three trends as significant in the past year.

On their own, each of these trends is a singular data point about how IT organizations are handling various challenges. Taken together, they suggest a larger story about how CIOs, CTOs and CDOs are modernizing their organizations, embracing AI experimentation, solving data problems and driving resource-stretching efficiencies—all necessary steps to meet the opportunities of advanced AI head-on.

TREND ONE:

PYTHON IS THE LANGUAGE OF CHOICE FOR AI PROGRAMMING

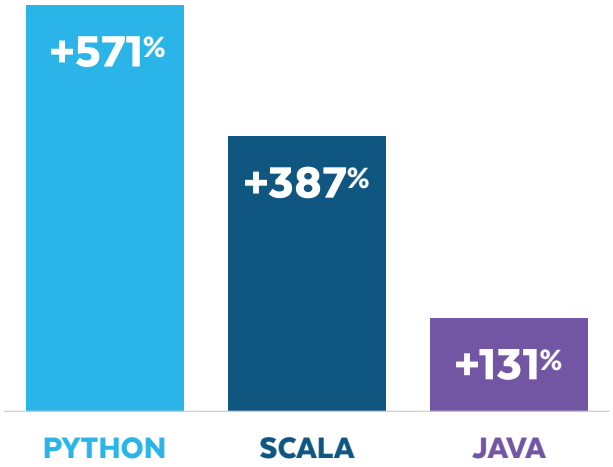
Developers are able to work with a variety of programming languages in Snowflake, and it's with interest that we note which languages are growing in popularity. In the past year, Python has surged.

Python has a lot going for it, including:

- **It's easy to learn and read, letting developers focus on solving AI problems rather than parsing abstract syntax.**
- **It has a vast ecosystem of libraries and frameworks that simplify potentially daunting AI tasks, from implementation of neural networks to natural language processing.**
- **It has a big, active community of contributors, which accelerates learning and problem-solving.**
- **It's flexible and portable, so developers can deploy AI applications across different platforms, systems and environments.**
- **Its extensive data-handling capabilities make it easy to manipulate data, which is a core challenge of any AI/ML project.**

Overall, Python lets devs focus on the problem, not the language. They can work fast, accelerating prototyping and experimentation—and therefore overall learning as dev teams make early forays into cutting-edge AI projects. And in the Snowflake Data Cloud, devs are seriously embracing Python.

In Snowpark, which expands programmability in Snowflake, Python use grew considerably faster than both Java and Scala in the last fiscal year: Python grew by 571%, while Scala grew by 387% and Java grew 131%.



AI-friendly Python significantly outpaced Scala and Java growth in the Data Cloud.

AI/ML IS GROWING WITH PYTHON

As Python use skyrockets in Snowpark, usage of some of the most popular AI/ML open source Python libraries in Snowpark has increased by 335%, including:

SCIKIT-LEARN IS UP
474%

XGBOOST IS UP
357%

Developers are bringing more AI/ML work to Snowflake, because they need a unified data platform and access to huge amounts of data used to build, train and run advanced models. But we believe the increase represents not only a shift of existing work to our platform, but a net increase in experimentation with advanced AI.

TREND TWO:

ENTERPRISES ARE FINALLY TAPPING THEIR UNSTRUCTURED DATA

Most data is unstructured, and most enterprises struggle to do much with it. This is not a problem that's going to go away. According to IDC, 90% of the data generated by organizations in 2022 was unstructured.[1]

Extracting value from that data has been a tech challenge for years, exacerbated by the near-simultaneous arrivals of smartphones and social media, and complicated by evolving regulatory regimes and privacy practices that govern all of an enterprise's data, structured or not. That last point is important; even as automation and artificial intelligence help us extract meaning from unstructured data, the actual management of it becomes more difficult.

Despite the challenges, Snowflake users are getting value out of unstructured data, especially with the growth of AI/ML. These data types are being processed with Python, Java and Scala, languages commonly used by data engineers, data scientists and app developers. The suite of languages for unstructured data processing became publicly available in public preview or general availability on June 27, 2023.

Given that Python in particular is the language of choice for many developers, data engineers and data scientists, its fast-growing adoption suggests that these unstructured data workflows are not just for building data pipelines, but also involve AI applications and ML models.

PROCESSING OF

UNSTRUCTURED DATA

+123%

FROM JULY 2023 TO JAN. 2024

1. IDC White Paper, sponsored by Box, "Untapped Value: What Every Executive Needs to Know About Unstructured Data," IDC #US51128223, Aug 2023

TREND THREE:

ENTERPRISES ARE MORE GRANULAR IN THEIR DATA GOVERNANCE

The last foundational trend is certainly not the least. Governance is absolutely essential to data strategy broadly, and AI strategy in particular. The outputs of LLMs and generative AI can be inaccurate or inappropriate, and a strong governance regime helps limit negative surprises.

In last year's trends report, we noted that with both data regulations and consumer privacy concerns on the rise, we had seen increased adoption of data governance features. In short, we saw that our users were applying more tags governing access and use of their data, meaning that they were ensuring that necessary audiences could make use of their data while restricting unauthorized user access. This year, that trend continues and in fact deepens.

We've seen significant increased adoption of governance features in a way that indicates not merely restriction, but control. The wide embrace of multiple governance features suggests that users want granular control over data to make it appropriately available to more users, for more use cases. This refined control is necessary to responsibly unlock the value of sensitive data.

Among the indicators of a more granular approach to data use, we saw use of the following governance features rise year over year:

- **The number of tags applied to an object rose 72%.**
- **The number of objects with a directly assigned tag is up almost 80%.**
- **The number of applied masking or row-access policies increased 98%.**
- **The number of columns with an assigned masking policy grew 97%.**
- **The cumulative number of queries run against policy-protected objects is up 142%.**

That last stat is particularly significant. There's a popular misconception that governance is about saying no, that it slows down or limits data innovation. While good governance is meant to put the brakes on genuinely unsafe or inappropriate activities, it's also an enabler of effective, responsible data usage. We're seeing more and more governance through the use of tags and masking policies, but the amount of work being done with this more carefully governed data is rising rapidly.

We expect these trends to continue as more and more enterprises improve how they govern their data, increase their responsible usage of it, and reap the benefits that data provides to their bottom line.

CUMULATIVE NUMBER OF JOBS RUN AGAINST POLICY-PROTECTED OBJECTS,

+142%

AI SCALES WITH APPS

While the establishment of a solid data platform and a strategy that breaks down silos and finds efficiencies has been a well-understood goal for years, AI is still mostly untapped by the enterprise. In the year that LLMs and generative AI have been in the media glare, many enterprises have begun to experiment, launching initial projects.

Within the Snowflake Data Cloud and the Streamlit community, we're able to measure activity in the LLM space and around application development, and throughout 2023 we saw great enthusiasm to get to work.

As with the foundational section, we've identified four trends in these early days of advanced AI.

A challenge of measuring trends in the enterprise AI space is that there's no precedent. In some cases, we made features available during 2023, so we don't have years of previous data to compare. What we have seen is enthusiastic uptake, and patterns of preference that we think point the way for these early days.

TREND ONE:

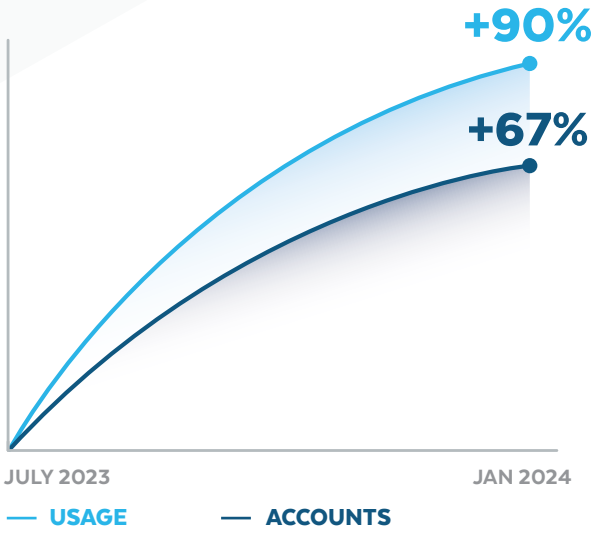
THE DEMOCRATIZATION OF AI IS HERE

A significant promise of LLMs and generative AI is that you don't have to be a highly trained data scientist to work with them. Natural language interfaces mean that you can talk to the data—or rather, the app that sits on top of the data—like a human, and the data/app will deliver its answers in a reasonable approximation of human conversation, too. That amounts to a "democratization of AI," as the tech marketers like to say. And it's here.

While this year's report does not have year-over-year statistics, what we saw in 2023 was tremendous, widespread enthusiasm. The fast adoption of the ML-based functions available in Snowflake Cortex shows how fast AI can happen when there is a solid foundation of data in place. These functions make it easier for those who aren't data scientists to work with machine learning algorithms.

- **The number of active accounts using ML-based functions[2] grew 67% between July 2023 (the first full month after public preview) and January 2024. That surge of initial growth, sustained over the remaining six months of the fiscal year, indicates the enthusiasm for, and the utility of, these "democratizing" functions.**
- **Comparing July 2023 to January 2024, monthly usage grew 90%.**

These are early days, and of course that growth surge starts from a relatively small initial point, but we're excited to see sustained and growing interest in tools that put more and more of the power of advanced AI into the hands of less-technical users. This frees the relatively small (and overwhelmed) teams of data scientists from being a bottleneck, and allows those experts to concentrate on the most complex and high-value projects.



Since ML-based functions became available in late June, more adoption by user accounts, and rising overall usage, indicate early steps toward the democratization of AI.

Note: Growth was not linear. This graphic illustrates the difference between the start and end points.

2. ML-based functions evaluated for this report include anomaly detection, forecasting and contribution explorer, which all went into public preview on June 27, 2023. Anomaly detection and forecasting were subsequently announced into general availability on Dec. 18, 2023.

TREND TWO:

THE LLM EXPLOSION IS HAPPENING NOW—PROBABLY AT YOUR OFFICE

When generative AI and LLMs became the singular topic of tech conversations a year and a half ago, we were assured that this technology would be everywhere, infiltrating every aspect of how we live and work. We can't say that this reality has fully materialized yet, but we're definitely seeing a lot of effort to get us there ASAP.

- **Within the Streamlit developer community, between April 27, 2023, and Jan. 31, 2024, we saw 20,076 unique developers work on 33,143 LLM-powered apps (this includes apps that are still in development).**
- **Historically, the Streamlit community has had a large percentage of non-corporate users, so we wondered if this massive surge might mostly be solo experimentation. But in a survey of 1,479 respondents, nearly 65% said their LLM projects were for work.**

And it seems that these developers are steadily improving their creations. Vector databases and vector search help improve the creativity and utility of an LLM app by making connections between related concepts rather than requiring exact word matches. The result is smarter, more accurate outputs, faster.

20,076

DEVS WORKED ON

33,143

LLM-POWERED APPS IN

9

MONTHS

TREND THREE:

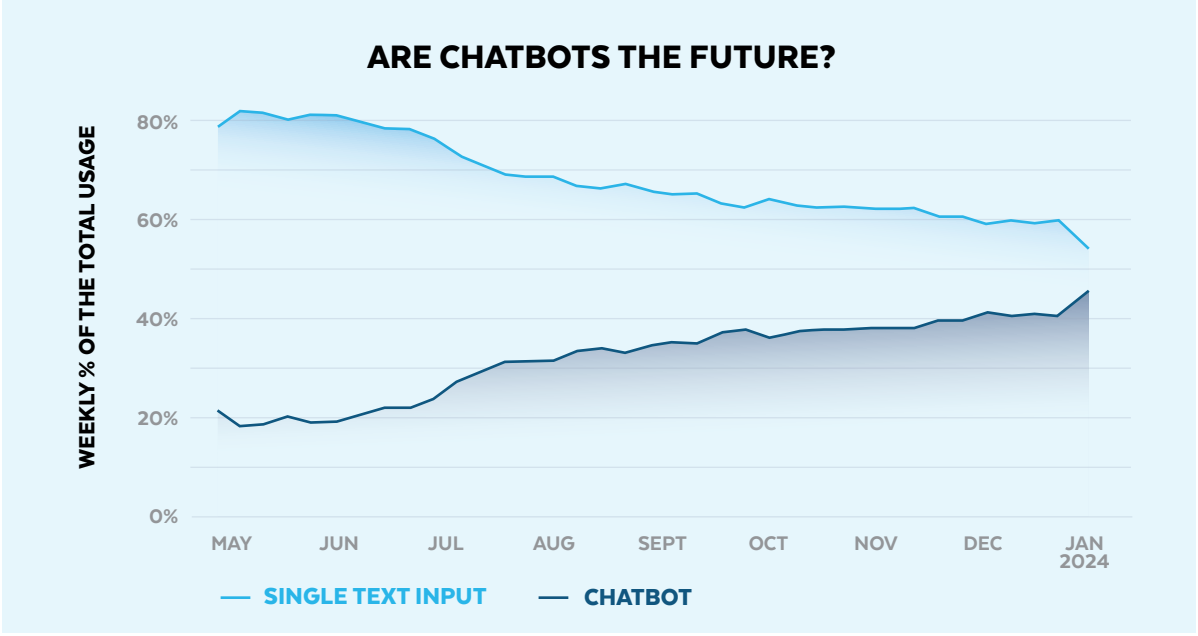
THE CHATBOT IS ON THE RISE

The great thing about a conversational interface is that you can have a conversation. We've seen in recent months a decided shift from the easy-to-build, straightforward single-text-input LLM toward the chatbot, which allows refinement through iterative text input.

Looking again at the more than 20,000 LLM-powered apps being developed with Streamlit, we see a definite direction for the chatbot, and it's up. In the week starting April 30, 2023, single-text-input apps peaked at 82% of all LLM apps built with Streamlit, leaving 18% for the chatbots.

From that point, the single-input line trended down and the chatbot line rose. By the end of January 2024, chatbots accounted for 46% of LLM apps, with single-input apps comprising 54%.

The steady climb of the chatbot probably does not represent a shift in the market's appetite for LLM apps. More likely, developers are increasingly able to make more complex chatbot apps to offer greater versatility and interactivity to meet both business needs and user expectations.



A SNAPSHOT OF DEV CONCERNS

In a community-wide survey, more than 980 Streamlit users selected their top concern, from a list of four common worries, about working with LLMs. The results were:

TRUST: Is the LLM response accurate?
36%

PRIVACY: Is my data safe?
28%

COST: AI ain't cheap!
19%

SKILLS: I'm still learning
17%

TREND FOUR:

ENTERPRISES WANT APPS AND DATA WITHIN A UNIFIED DATA PLATFORM FOR BETTER SECURITY AND GOVERNANCE

You don't have to build your LLM application on the same platform as your data, but there are significant advantages to doing so. By having unified data governance and not having to move data across compute environments, application development is faster, deployment is easier, and operational maintenance costs are lower.

Therefore, to continue practicing what we preach about bringing the work to the data, rather than vice versa, we introduced the Snowflake Native App Framework in 2023.

Snowflake Native Apps let users deploy applications within the Data Cloud, leveraging the Snowflake platform to run all three layers of the app, including data, processing and user interface. But the question is, does anyone actually want that?

The early answer appears to be "Yes." The Native App Framework went into public preview on June 27, 2023. Comparing July 2023 to January 2024:

- **We've seen 311% growth in the number of Snowflake Native Apps published.**
- **We saw 147% growth in installation/adoption of these applications.**
- **Usage of these apps grew 96%.**

What this means is that, given the choice, users want to build applications within their data platform—where the data is—rather than export copies of the data to external technologies.

And frankly, it makes sense. We've seen that a strong data foundation prepares an organization to succeed with AI. That enterprises would want to work within a solid data platform to create their applications is an extension of that principle. We believe this will soon be an industry-wide baseline.

CYBERSECURITY WORK IS MIGRATING TO THE DATA PLATFORM

Underscoring the trend to bring work to the data, we're seeing a rise of cybersecurity workloads being brought to the Snowflake Data Cloud.

- For cybersecurity connected apps, where a SaaS vendor stores and processes data in the end consumer's Snowflake instance, the average number of connected accounts **increased 72%** year over year.

This tells us that cybersecurity teams see the value of doing security work within their company's unified data platform, rather than through externally managed applications.

FROM FOUNDATION TO ELEVATION

IT teams are used to how much work occurs on the backend to provide a positive, painless experience. The simplest application hides a lot of complexity. That's definitely true with LLMs and generative AI. We're seeing that organizations understand this and are fortifying their data foundation even as they make their first forays into cutting-edge AI.

Some of the foundational trends we're seeing apply directly to AI: robust, refined governance; increased use of Python; coming to grips with the vast quantities of unstructured data. Others speak to a general excellence and willingness to adopt new practices to accelerate time to value, such as the growth of serverless computing.

As organizations progressively improve their foundation, they pave the way for successful AI initiatives that will deliver reliable, ethical, secure and impactful results. And the trends we're seeing in the AI and applications spaces suggest progress is being made.

Organizations are picking their models, creating more complex LLM applications, making AI more available to a wider range of users, and reaping the benefits of a unified data platform. There has been a lot of hype around the transformational potential of AI, but judging from what we're seeing in the Data Cloud, the frenzied fanfare is beginning to materialize into concrete results.

NEXT STEPS

Learn more about how Snowflake can help you improve your data foundation and launch successful AI initiatives.

SNOWFLAKE FOR AI AND ML

See how you can securely build and deploy LLMs and ML models in the Data Cloud.

LEARN MORE

SNOWPARK

Runtimes and libraries that securely deploy and process Python and other programming languages in Snowflake.

LEARN MORE

SNOWFLAKE HORIZON

Snowflake's built-in governance solution provides a unified set of compliance, security, privacy, interoperability and access capabilities in the Data Cloud.

LEARN MORE

STREAMLIT IN SNOWFLAKE

Turn data and ML models into interactive apps with Python—now all in Snowflake.

LEARN MORE

APPENDIX: METHODOLOGY

The Snowflake Data Trends Report 2024 is generated from fully aggregated, anonymized data detailing usage of the Snowflake Data Cloud and its integrated features and tools. In this report, we examine patterns and trends in data and AI adoption across more than 9,000 global Snowflake accounts. The Snowflake Data Cloud provides insight into the state of data and AI, including which technologies are the fastest growing. Note that usage attributable to internal consumption, if any, has been removed and is not reflected in any of the metrics contained herein. The accounts and usage reflected in this report represent every major industry and include both longtime Snowflake users and others who only recently joined the Data Cloud.

Except where noted in the text, the data in this report compares monthly averages from January 2024 (represented as "this year") to averages in January 2023 ("last year"). When compared, this is depicted as "year over year" growth to align with Snowflake's fiscal year end, though the figures themselves are only representative of January figures to calculate growth.

When possible, we have provided these year-over-year comparisons to showcase growth trends over time. Where data was drawn from Snowflake features that became publicly available after the start of the fiscal year, data was collected and compared as of the first full month after which the feature became available in public preview, and that date is noted in the text. Notably, growth figures for features moving into public preview are expected to be considerably higher, as private previews are limited in scope and necessarily restricted to select Snowflake customers.

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 artificial intelligence (AI) / machine learning (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 the Snowflake Data Cloud to power their businesses.

Learn more at snowflake.com