snowflake®

DATA TRENDS 2024
HEALTHCARE AND LIFE SCIENCES

How Industry Leaders Are Building for Success in the Snowflake Data Cloud

TABLE OF CONTENTS

GEN AI IS
TRANSFORMATIONAL

In 2023, the hype surrounding generative AI was massive. But in 2024, the real work is underway, as enterprises start to pour their proprietary data into large language models (LLMs) and create bespoke applications that drive their businesses in new directions.

At Snowflake, we see this trend clearly, with some of the largest enterprises in the Global 2000. Gen AI has moved out of the experimentation phase and into production. Our recent report, **Snowflake Data Trends 2024**, reveals how organizations have been using this transformative technology and other data-driven workloads within the Snowflake Data Cloud from January 2023 to January 2024. Here are the topline results across all industries:

- **The number of active accounts adopting machine learning (ML) functionality within Snowflake increased 67% since the machine learning functions of Snowflake Cortex went into public preview in June 2023.**

- **Organizations' usage of data governance measures is rising, and seems to be improving their ability to use data: While usage of tagging data increased 70%–100%, the number of queries against protected objects is up 142%.**

- **The number of native apps created by Snowflake users increased by 311%, while adoption of these apps soared 96%.**

- **Usage of Python, a very popular language for AI development, grew 571% — considerably more than any other language year over year.**

In this report, we'll take a closer look at how Snowflake users are working with their data in healthcare and life sciences to simplify and solidify their data foundation and explore the factors influencing AI adoption — and hesitation — in healthcare.

METHODOLOGY

We looked at how our Snowflake healthcare and life sciences users 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. Where relevant, we compared industry usage to our broader, cross-industry metrics to show both alignment and, more importantly, deviations. Generally, we compared usage in January 2023 to January 2024 to align with Snowflake's fiscal year, except in cases where features became 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.**

GEN AI FOR HEALTHCARE AND LIFE SCIENCES:
FIRST, TAKE A LONG, HARD LOOK AT YOUR DATA FOUNDATION

With sensitive patient information at risk, data security will always be a top concern for healthcare and life sciences. So when we broach the subject of AI in this industry, the question that precedes everything else is: How do we implement a responsible AI strategy?

In July 2023, the U.S. Department of Health and Human Services (HHS), in coordination with its Health Sector Cybersecurity Coordination Center, published a **threat briefing** focused on AI, cybersecurity and the healthcare industry. It focused largely on bad actors executing attacks along the lines of phishing emails but set off wide-reaching alarm bells for security personnel. The HHS explicitly recommended penetration testing, ongoing threat monitoring, cyber threat analysis and incident handling, along with AI training for cybersecurity personnel.

In October 2023, the White House **issued an executive order** instructing the HHS to create an AI Task Force by January 2024. The task force was given 365 days from launch to develop a regulatory action plan for predictive and generative AI-enabled tech in healthcare and life sciences. That plan is meant to cover a gamut of aspects — from AI in healthcare delivery and financing to mandating that model documentation is made available to users so they can confirm whether AI is being used in a safe, governed way.

The heightened fear of data breaches at the expense of AI makes effective data governance more crucial than ever. And that is consistent with our findings in our **Data Trends 2024 Report**.

FIVE CONCERNS SURROUNDING GEN AI

Generative AI's creative capabilities are groundbreaking. This transformative technology is likely to have a profound impact across all sectors of the economy, especially in healthcare and life sciences. But gen AI advances also come with **a number of potentially serious risks**. Here are the most significant areas of concern:

1. **Unintentional bias:** ML models' training data may be biased. For example, over-sampling of a particular population is shown to skew the results of facial recognition algorithms, adversely impacting some demographic groups.

2. **Lack of transparency:** In many cases, the training for a particular model is unknown or undisclosed. This makes it difficult to determine whether the data is based on unreliable sources or is otherwise biased. In addition, the neural networks used to build LLMs are typically "black box" applications whose inner workings are opaque even to the people who designed them. The ability to demonstrate how an AI model makes decisions is likely to be addressed by future regulatory activity.

3. **Data leaks:** Public-facing gen AI chatbots can leak source code, or personal or proprietary information. These bots may also inadvertently reproduce materials used in their training data — a flaw known as "regurgitation."

4. **Inaccurate or false results:** AI chatbots sometimes generate fictional information in response to queries — commonly called "hallucinations," so results cannot be taken at face value.

5. **Cybersecurity issues:** Bad actors use AI to corrupt the AI model's training data (known as "data poisoning") or to extract sensitive information from chatbots using carefully constructed prompts (known as "prompt injection attacks").

Even the most optimistic healthcare and life sciences technologists acknowledge that most companies in the industry need to take a hard look at their data foundations before they can leverage AI to its fullest potential. There are opportunities to vastly improve claims processing and health equity, optimize clinical development, modernize supply chains, accelerate drug discovery, and more — but cementing the data foundation needs to come first.

Since the HHS AI Task Force has another year to establish guidelines for the industry, it's no surprise that security and compliance teams are hesitant to act now. In some cases they're stopping AI use cases before they even get started. Responsible AI is built on a strong data foundation, and that likely requires cloud-first infrastructure and organization wide data literacy. For organizations that already have a solid data foundation, there are a couple of steps to take to try and avoid the big red stamp of denial. You should assess your organizational data maturity then establish an AI committee with representation from internal stakeholders, such as privacy, procurement, legal and IT teams before even beginning to consider AI use cases.

It is crucial to provide extensive background on intended use cases and data security maintenance, as well as clarity on whether data will be moved for AI use cases to the security team from the get-go.

.

TRENDS THAT MATTER
TO HEALTHCARE AND LIFE SCIENCES

Healthcare and life sciences organizations will have to reconcile the industry's slower shift to cloud-first data strategies to make AI a viable option. In recent years, there's been an industrywide recognition that multi-cloud approaches can support cost optimization, improve resilience, provide enhanced security controls to ensure compliance with applicable regulations, and allow for the adoption of innovative solutions.

Our research identified four significant trends that show how data modernization is having a foundational impact on healthcare and life sciences organizations — influencing everything from organizational strategy to planning operations and technology investments.
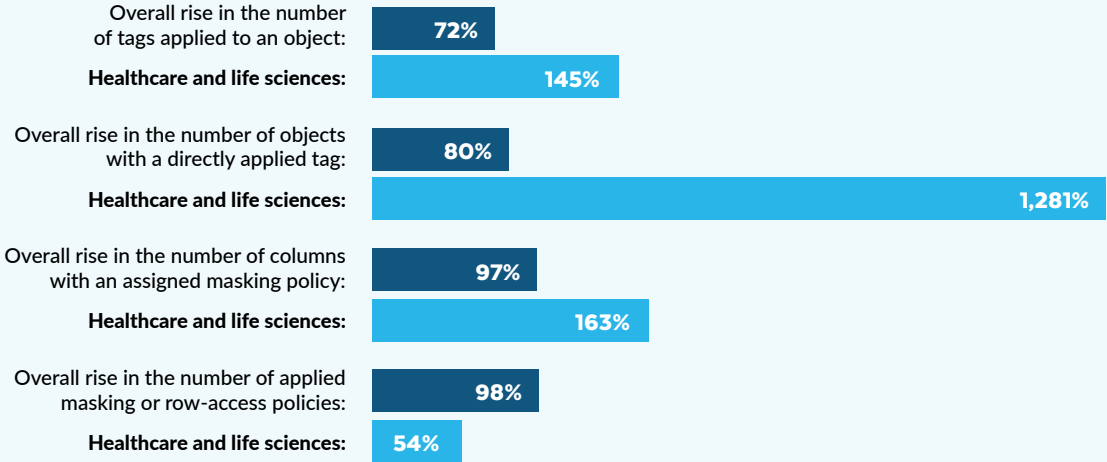
TREND ONE:
DATA GOVERNANCE IS THE FOCAL POINT

Healthcare and life sciences is a heavily regulated sector. Data quality and accuracy directly impact patient care and safety, making data governance a top concern. Data governance efforts are rising across the board, with the number of queries against policy-protected data assets increasing 142%. Notably, healthcare and life sciences saw a smaller jump here, at 76%. The illuminating context is that data and security regulations, like HIPAA in the U.S., impose extremely strict rules around data access and handling, which drives organizations to embrace a more conservative approach to querying policy-protected data.

In general, data governance measures rose 70%–100% in the last year. Healthcare and life sciences organizations took full advantage of tagging — which is critical to protecting and masking PHI and PII data, including the 18 identifiers outlined in the Safe Harbor provision of HIPAA, across thousands of tables and tens of thousands of columns — with the number of tags jumping 145%, twice the overall rate, and the number of objects with a directly applied tag growing by 1,281%. This rise suggests that well-governed data is being used more, not less.

As the graphic here shows, healthcare and life sciences organizations rely more heavily on tagging than masking, and came in below the general growth rate in terms of applied masking or row-access policies. However this industry's increase in the number of objects with a directly applied tag was significantly over the average. This is most likely a result of previous governance efforts and well-established access policies, spurred by Federal Communications Commission compliance requirements, including the HIPAA and the HITECH Act regulations.

ADOPTION OF GOVERNANCE FEATURES IS ON THE RISE

Overall rise in the number of tags applied to an object: **72%**

Healthcare and life sciences: **145%**

Overall rise in the number of objects with a directly applied tag: **80%**

Healthcare and life sciences: **1,281%**

Overall rise in the number of columns with an assigned masking policy: **97%**

Healthcare and life sciences: **163%**

Overall rise in the number of applied masking or row-access policies: **98%**

Healthcare and life sciences: **54%**

TAGGING — WHICH POWERS CLASSIFICATION, SENSITIVE DATA TRACKING, RESOURCE USAGE MONITORING AND DATA PROTECTION — JUMPED

145%

THAT'S TWICE THE OVERALL RATE

TREND TWO:

ENTERPRISES ARE TAPPING INTO UNSTRUCTURED DATA

According to IDC, 90% of the data generated by organizations in 2022 was unstructured.[1] The healthcare and life sciences industry houses high volumes of unstructured data, including doctor's notes, discharge summaries, post-encounter notes, audio recordings from speech therapy or patient narratives from clinical trials, and medical imaging such as MRIs.

The processing of unstructured data increased 123% across the Snowflake Data Cloud since the features in Snowpark became publicly available on June 27, 2023. But within the healthcare and life sciences industry, that figure was 97% — less than the overall rate.

While 90% of all data in the world is unstructured, the focus in healthcare and life sciences is getting value from the overwhelming amounts of structured data — like electronic health records, claims data, patient demographics, lab results, vital signs and more — making up the remaining 10%. The sophistication level of optimizing structured data varies across the industry, with payers and providers and life sciences having a more mature approach, while healthcare providers overwhelmingly continue more foundational digitization efforts.

Given the high regulatory requirements and the industry's continued modernization of more structured data, organizations seem to be taking a slower approach toward working with unstructured data.

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

TREND THREE:

PYTHON USAGE IS UP DRAMATICALLY

The Python programming language is particularly well-suited to building AI apps. So, as organizations began to strategize and experiment around LLMs and generative AI, it was no surprise that we started to see considerably more work in Python. Over the 12 months between January 2023 and January 2024, Python usage grew much faster than any other scripting language in Snowpark, our coding library component of the Snowflake Data Cloud.

Across all industries, Python use grew 571% in the last fiscal year, easily outpacing Scala (up 387%) and Java (up 131%). For healthcare and life sciences organizations, the use of Python increased by 427% in the same period.

Any increase in Python use as a programming tool indicates that healthcare and life sciences organizations are performing more advanced processing of data using Snowpark — jobs that would have been processed on premises or using Spark.

Healthcare data is inherently disorganized, inconsistently mapped, and needs to be prepped for action. Python's open-source roots, explainability and greater accessibility make it a better choice for some organizations than more statistically-based programming languages, such as R, MATLAB or SAS.

These findings suggest that IT teams in the healthcare and life sciences field are in a good place to launch advanced AI initiatives since they've already oriented their teams around Python.

TREND FOUR:

THE DEMOCRATIZATION OF AI

One of the breakthrough benefits of generative AI is that it can be quickly deployed and used by team members with little to no data science expertise. That removes potential bottlenecks in making data available for analysis, and increases the ability of business leaders to take advantage of this transformative technology to drive business decisions.

At Snowflake, we've seen this trend across all industries. In June 2023, we debuted the machine learning functions of Snowflake Cortex in public preview. Cortex is a fully managed service that enables organizations to quickly analyze data and build AI applications within the Snowflake Data Cloud. Starting from the first full month of public preview in July 2023 through January 2024, the number of active accounts adopting of ML-based functions within Snowflake Cortex grew 67% overall.

Healthcare and life sciences companies are moving ML inference into Snowflake for use cases such as applying predictive analytics to identify high-risk factors, improving personalized care plans, and accelerating drug discovery and development.

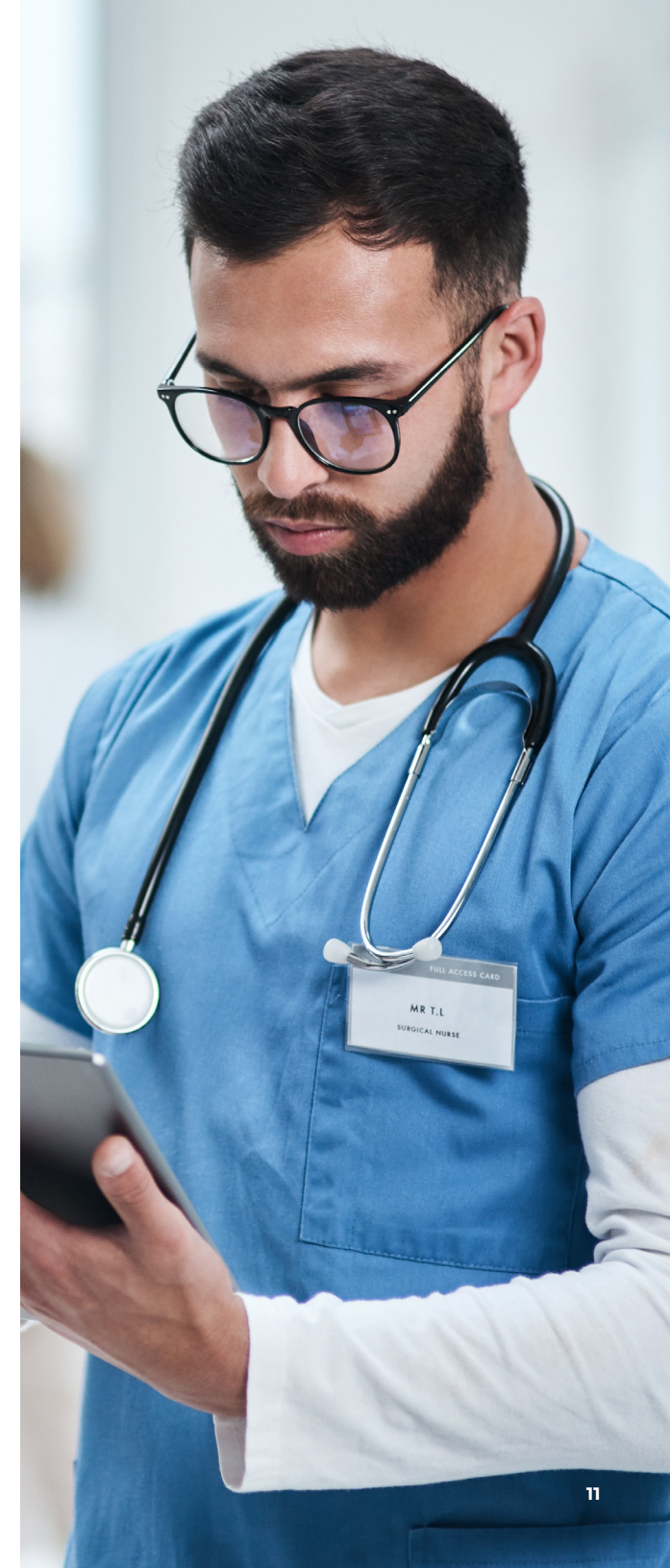# THE STATE OF THE HEALTHCARE AND LIFE SCIENCES DATA CLOUD

The ability of healthcare and life sciences organizations to collect, share and analyze data will be one of the primary factors determining their ongoing success. Data is the key to controlling costs, improving patient outcomes, ensuring data compliance and security, enhancing operational efficiency, and enabling real-time analytics capabilities.

Simply collecting data is not enough. To get the benefits of AI and predictive analytics, healthcare and life sciences organizations will need to modernize their data, break down internal data silos, and adopt a unified platform that lets them share data across their organizations and with external partners in a secure and compliant way.

A unified, cloud-based data platform like the **Healthcare and Life Sciences Data Cloud** can help healthcare and life sciences organizations use data effectively for better decision-making, operational excellence and cost efficiency.

In a rapidly transforming business environment, the Snowflake Healthcare and Life Sciences Data Cloud offers several key advantages:

- **Wrangle the explosion of healthcare and life sciences data:** According to **RBC Capital Markets**, healthcare data will make up 36% of the world's data by 2025, outpacing manufacturing, financial services, and media and entertainment. The Data Cloud helps healthcare and life sciences organizations prepare for the ongoing onslaught of data by simplifying their architecture, centralizing data, providing governance capabilities and accelerating innovation.

- **Achieve a unified view of enterprise data:** Snowflake is a single, easy-to-use global platform that allows healthcare and life sciences enterprises to do more with less. It's accelerating how hospitals, insurers, biotech and other ancillary organizations simplify access to data and enable secure data collaboration. Healthcare and life sciences organizations can connect with an ecosystem of top data and app providers, prominent partners and critical service providers. It alleviates operational and IT burdens to increase agility and cost savings, so enterprises can focus on development.
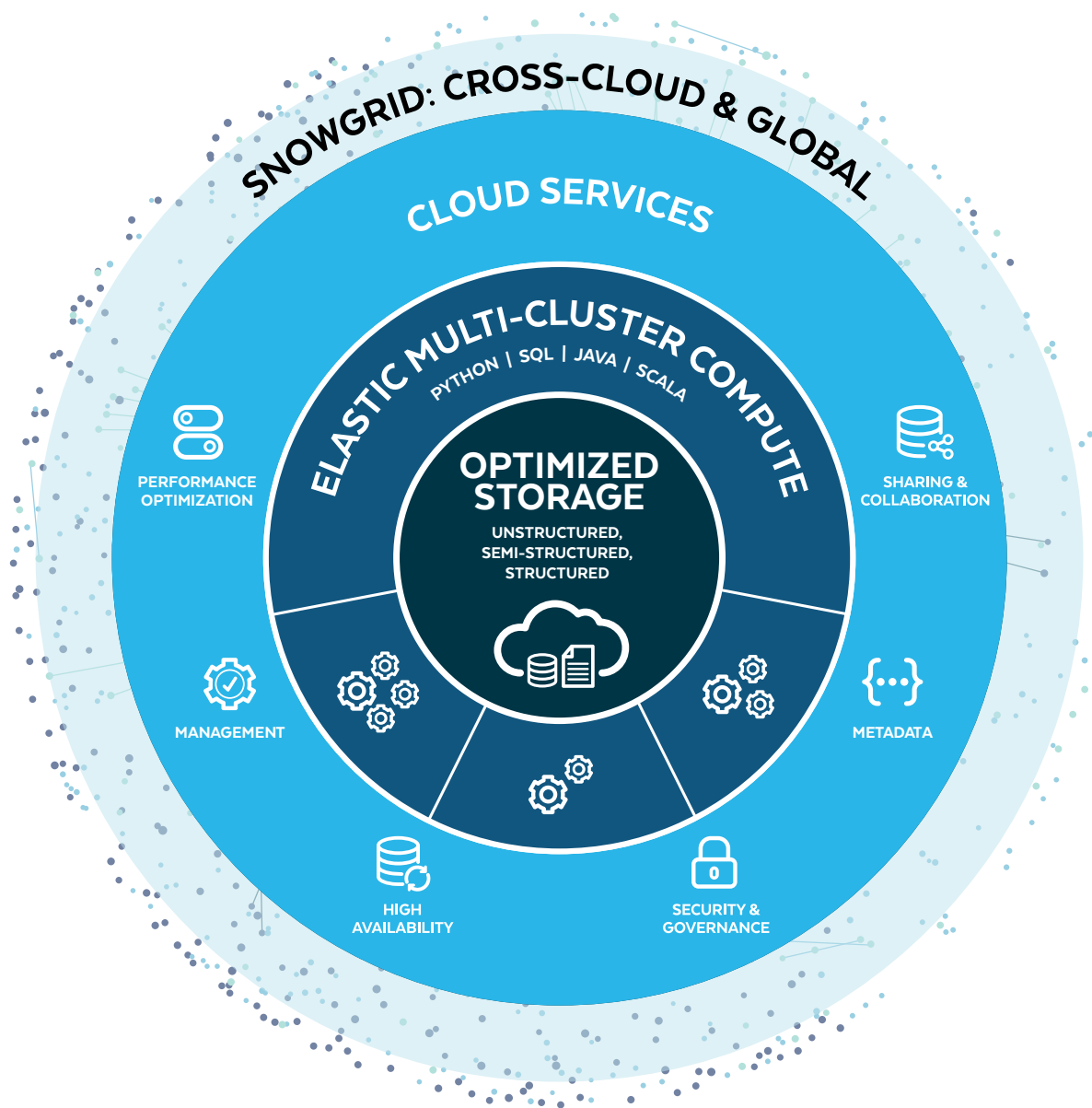
Snowflake Platform Architecture

- **Benefit from the potential of AI:** With the Healthcare and Life Sciences Data Cloud, companies can deploy AI solutions to improve patient outcomes, regulatory compliance and cost pressures; expedite clinical processes; and improve decision-making and competitiveness. Healthcare and Life Sciences organizations can customize their gen AI and LLM solutions in hours, easily and securely use gen AI and LLMs in seconds, and keep them in-house to secure and govern their data and IP. Generative AI offers an opportunity to democratize the benefits of AI and its insights beyond the AI experts.

- **Access data from all corners of the business and trusted partners:** The data sharing provided in the Snowflake Data Cloud is more efficient and secure than legacy methods such as FTP, SFTP and sometimes even email. This means data exchange — encompassing data models, research, performance analytics, risk analytics and patient insights — happens more quickly and at a lower cost. And since such transactions can happen thousands of times a day and involve petabytes of data, even efficiencies measured in milliseconds create significant advantages.

USING DATA SHARING TO IMPROVE PATIENT OUTCOMES AND SOLVE CLINICAL AND BUSINESS CHALLENGES

The Snowflake Data Cloud modernizes how organizations access, share and leverage data. With Snowflake Marketplace, healthcare and life sciences companies can access live, ready-to-use data and services. Here are three important ways organizations are using data sharing and collaboration to solve critical business problems:

Sourcing real-world evidence

HCLS organizations have direct access to traditional patient information like pharmacy and medical claims, reimbursements, electronic healthcare records (EHRs), and visitation data. However, according to the World Health Organization, studies suggest that social determinants of health data — including income, housing, education, job and food security — account for **up to 55%** of health outcomes. Global, real-world evidence can help healthcare and life sciences organizations bridge the gap between treatment effectiveness and patient lifestyle.

Here are some Snowflake Marketplace providers that healthcare and life sciences companies use to source real-world evidence:

- **IQVIA:** IQVIA is a leading global provider of advanced analytics, technology solutions and clinical research services to the life sciences industry. Dedicated to delivering actionable insights, IQVIA has the largest global healthcare data network in the world, with over 1 million data sources providing access to more than 800 million nonidentified patient records and over 95 billion healthcare records processed annually in over 100 countries.

- **Definitive Healthcare:** Definitive Healthcare's passion is to transform data, analytics and expertise into healthcare commercial intelligence. They help clients uncover the right markets, opportunities and people to shape tomorrow's healthcare industry. Their SaaS platform creates new paths to commercial success in the healthcare market, so companies can identify where to go next.

Improving patient experience and outcomes

There's a massive demand for data sharing between payers and providers in healthcare and life sciences. Data sharing helps healthcare organizations adopt a human-centric approach that focuses on patients' needs and preferences, and streamlines business processes. By sharing data across hospitals, pharmacies, clinics, insurers, telehealth visits and wellness apps, healthcare organizations can create a comprehensive view of patient health. This holistic view helps address common inequities and allows for better care delivery, higher patient satisfaction and potentially better health outcomes.

Here are some Snowflake Marketplace providers that healthcare and life sciences companies can use to help power human-centric care:

- **NTT DATA:** NTT DATA is a $30 billion trusted global innovator of IT and business services. They help clients transform through business and technology consulting, industry and digital solutions, applications development and management, managed edge-to-cloud infrastructure services, BPO, systems integration, and global data centers. NTT DATA is committed to clients' long-term success and combine global reach with local client service in over 80 countries.

- **Integra Connect:** The Integra Connect platform enables providers, researchers and payers to harness clinical and claims data from a variety of sources and geographies for their decision-making.

Optimizing supply chain

Real-time data sharing enables companies to respond to changes in demand, optimize production, improve order fulfillment and identify potential disruptions within the supply chain network. Across the industry, improper categorization, the fast pace of patient care and other factors create huge inventory management challenges. Something seemingly insignificant, like overordering gloves, can cost organizations millions per year. One of the learnings from the massive global pandemic was that some vaccine distribution problems affecting patients could have been avoided with better supply chain visibility.

Here are some Snowflake Marketplace providers that offer improved supply chain transparency:

- **Resilinc:** Resilinc is the leading provider of comprehensive end-to-end supply chain resiliency solutions.

- **AccuWeather Data Suite:** The AccuWeather Data Suite combines weather data, technology and human insight to improve lives and businesses.

- **FactSet:** FactSet is an industry leader in acquiring, integrating and managing financial data. Its content gives customers the power to monitor the global markets, research public and private companies, and gain industry-level insight. FactSet's proprietary symbology links and aggregates a variety of content sources to ensure consistency, transparency and data integrity across your business.

FIVE WAYS HEALTHCARE AND LIFE SCIENCES ORGANIZATIONS CAN LEVERAGE GENERATIVE AI

Though it's still ramping up, generative AI already promises a wide range of benefits and potential use cases for healthcare and life sciences organizations. Here are five ways these organizations can benefit from this transformative technology:

1. Enhanced cost savings and efficiency

AI in healthcare is projected to save the **industry up to $150 billion annually by 2026**, primarily through enhanced operational efficiency. Gen AI can streamline administrative processes and optimize workflows — saving employees time and saving the organization money. It can improve efficiency in matching patients to clinical trials, claims processing and prior authorizations. On the clinical side, it can enhance EHRs by improving data accuracy and automating documentation processes.

2. Improved productivity in clinical applications

Generative AI can be used to improve clinical practice operations and patient care delivery. It has the potential to automate medical documentation, analyze unstructured data such as clinical notes and diagnostic imaging, and provide evidence-based recommendations — ultimately improving patient care and health outcomes.

3. Faster drug discovery and development

Generative AI can accelerate drug discovery by expediting research, designing novel compounds and optimizing trial designs, leading to faster patient recruitment and reduced trial durations.

4. Better patient experience and outcome

Gen AI can analyze massive amounts of patient data. The data insights it provides can inform algorithms designed to detect health risks and potential complications through proactive intervention, prevent adverse health outcomes, or aid in the development of personalized and effective treatment plans. Because AI-enabled devices can monitor patients remotely, providers are able to track symptoms in near real-time for more timely interventions, which can improve patient outcomes.

5. Assisted regulatory compliance

Generative AI can aid in regulatory compliance by navigating uncharted terrain to ensure that innovative solutions meet rigorous safety and ethical standards. Organizations can implement shared responsibility models, access compliance support services, bring the work to their data for security measures, and facilitate adherence to various state and country regulations and standards.

PREPARING FOR
THE AI FUTURE

Throughout the last fiscal year in the Streamlit developer community, we saw 20,076 unique developers work on 33,143 LLM-powered apps (including apps in development). This is a clear sign that a world filled with the power of AI is underway. It may not be evenly distributed yet, but it's here.

Our research strongly suggests 2024 will be the year that generative AI becomes a crucial component of the enterprise technology stack, and that gen AI applications may spur better data-driven decision-making. It also shows that unified data management and governance will be critical to making those efforts a success, both for healthcare and life sciences companies and the greater business community.

Today's forward-thinking organizations are creating more complex LLM applications, making AI more available across the enterprise, and seeing the benefits of a unified data platform. Looking into the Snowflake Data Cloud, we are beginning to see the power of AI transform into real business benefits.

Learn how the Data Cloud can help you build a strong data foundation and prepare for what's ahead.

APPENDIX: METHODOLOGY

The Snowflake Data Trends 2024: Healthcare and Life Sciences ebook 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 Snowflake's healthcare and life sciences accounts. These trends provide insight into the state of data and AI, including which technologies are the fastest-growing within the healthcare and life sciences space. 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 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 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