



REPORT

2024 State of Apache Airflow™

Presented by **ASTRONOMER**

Executive Summary

In an era characterized by the convergence of data and digital transformation, we present the inaugural **2024 State of Apache Airflow**—the first comprehensive analysis of the growth, usage, and user community dynamics of what has become a critical tool for organizations navigating the complexities of data pipelines development, data orchestration, and data delivery.

By now, the importance of data in driving business success and delivering value to customers and stakeholders is undeniable. Modern enterprises must now function as data-driven entities, with data orchestration comprising its central nervous system. The effectiveness with which data is orchestrated across cross-functional teams determines the business' ability to adapt to meet digital challenges and opportunities.

This reality is further underscored with the emergence of artificial intelligence (AI) and machine learning (ML) innovations, which have imposed increased demands for scalability and operational excellence across disparate teams. It's no wonder then that Apache Airflow—so well-suited to the challenges of scaling, data orchestration, and facilitation of best practices—has seen such a dramatic emergence as a dominant workflow management platform in the last year, especially for AI pipeline delivery →

The report highlights three critical trends we uncovered:

1.

Rapid growth in Airflow adoption

In the last year, Airflow downloads surged by two-thirds compared to the previous year, accompanied by the substantial growth in the open-source community and an overwhelmingly strong willingness among users to recommend it to others.

These metrics collectively reflect Apache Airflow's rising prominence as a proven, dependable, and continuously improving platform that meets the needs of organizations moving towards more agile and scalable data solutions. Users' positive experiences and endorsements further solidify its position as a dependable choice for data orchestration and workflow management.

2.

Crucial role in mission-critical data delivery

The data shows Apache Airflow's pivotal role in data delivery, with nearly half of users rating it as "very important" to their business. This underscores its growing recognition for reliability and flexibility, making it a cornerstone in data management and workflow orchestration.

3.

Standard for ML/AI pipeline development and delivery

A dramatic spike in AI and ML usage on Apache Airflow over the past year shows that a large number of users are adopting the tool, specifically for the advantages it offers in supporting AI and ML innovation. This trend positions Airflow as the go-to platform for AI development and deployment moving forward.

We invite you to read on to see how these three trends are playing out and how you can better position your organization to benefit from these trends. Undoubtedly, Apache Airflow stands at the forefront, empowering more organizations to maximize the potential of their data, accelerate their digital transformation journeys, and drive innovations in AI and ML with unparalleled efficacy →



Trend 1: Apache Airflow Experienced Exponential Growth

In summary, 2023 marked a watershed year for Apache Airflow. It experienced exponential growth in downloads, fostered a thriving community of contributors, garnered impressive GitHub stars, and most importantly, earned the trust and advocacy of its users. This makes it clear that Apache Airflow is not merely a rising star but an indispensable asset for both organizations and its developers →

 **67%**

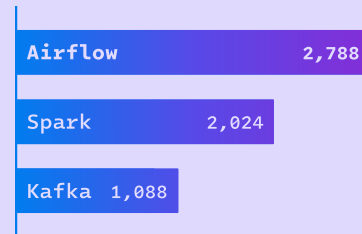
year-over-year increase in number of Airflow downloads with over 165.7 million¹



This exponential growth signifies the platform's appeal and its essential role in addressing the complex data orchestration needs of organizations.

 **2.8k**

Apache Airflow contributors outpacing stalwarts Apache Spark and Apache Kafka²



This robust community support ensures that Apache Airflow continues to evolve, adapt, and meet the evolving needs of users.

 **33k**

GitHub stars³

This acknowledgment further reinforces Airflow's significance in the open-source ecosystem and the trust placed in it by developers worldwide.

 **92%**

of users would recommend Airflow⁴

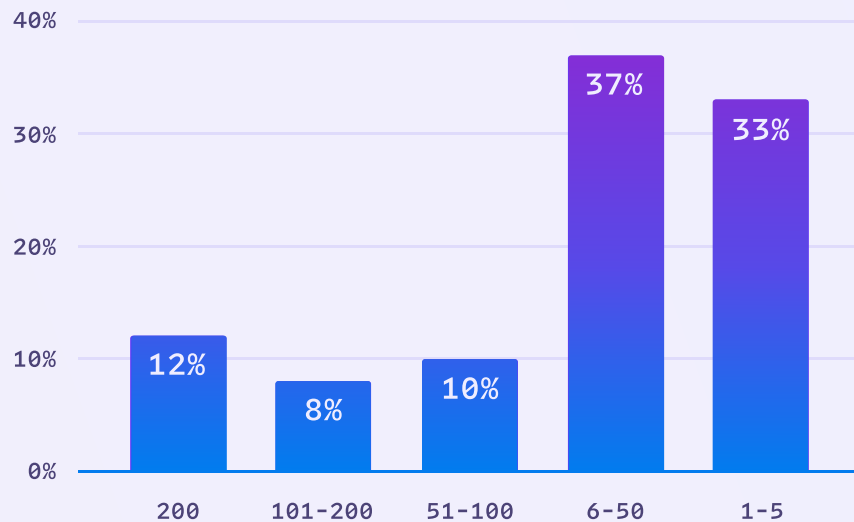
It's not just numbers and stars; it's also about the community that stands behind Apache Airflow. When asked about their willingness to recommend Airflow, 92% of respondents expressed their likelihood to do so.



Trend 2: Companies Rely on Apache Airflow for Mission-Critical Data Delivery

From widespread adoption to its central position in daily workflows, and its criticality in ensuring uninterrupted data delivery, the data reveals that Apache Airflow has evolved from a niche tool to a mission-critical asset pivotal role in modern data operations. Yet, as organizations increasingly tie data delivery to business-critical operations, Apache Airflow's prominence continues to grow, with teams expanding their use cases and exhibiting growing confidence in its capabilities.

People in company using Airflow



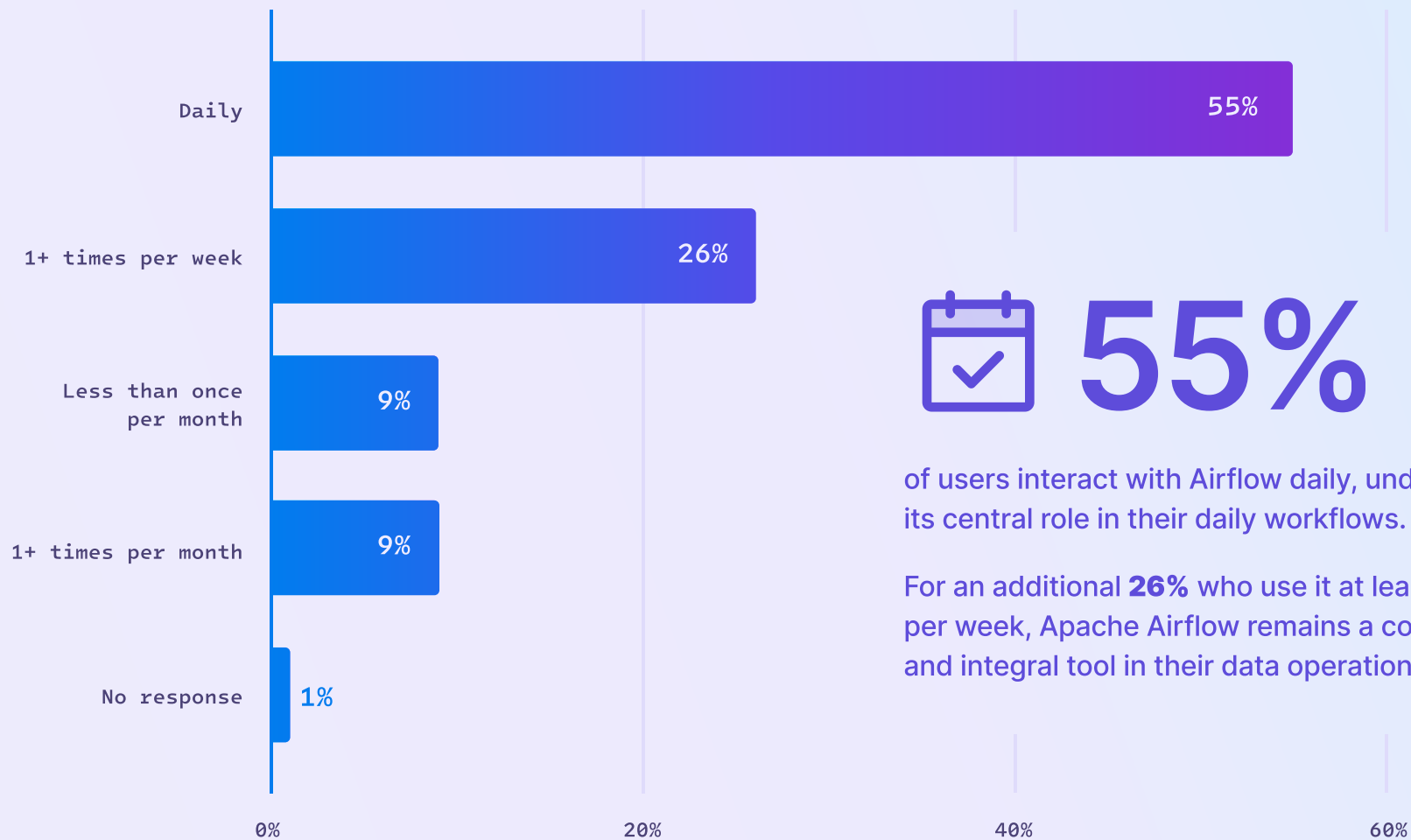
Source: 2023 Apache Airflow Survey, n=799



67%

of companies have more than **6** people using Airflow, highlighting the growing need for orchestration as data teams are increasingly tied towards delivering business outcomes.

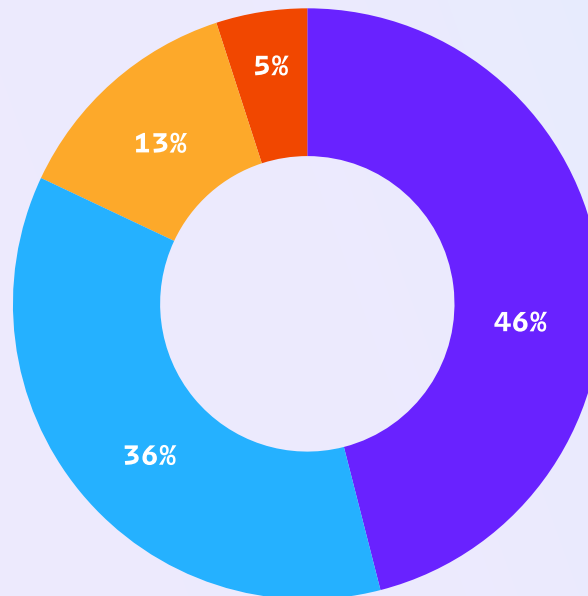
Frequency of Interaction with Airflow



Source: 2023 Apache Airflow Survey, n=799

How important is Airflow to your business?

- Very important. If it's broken, there is a severe impact across the entire business.
- Important. If it's broken, there is an impact on a couple of teams.
- Important, but limited in scope to one team.
- Not very important. It is not responsible for business critical workflows.



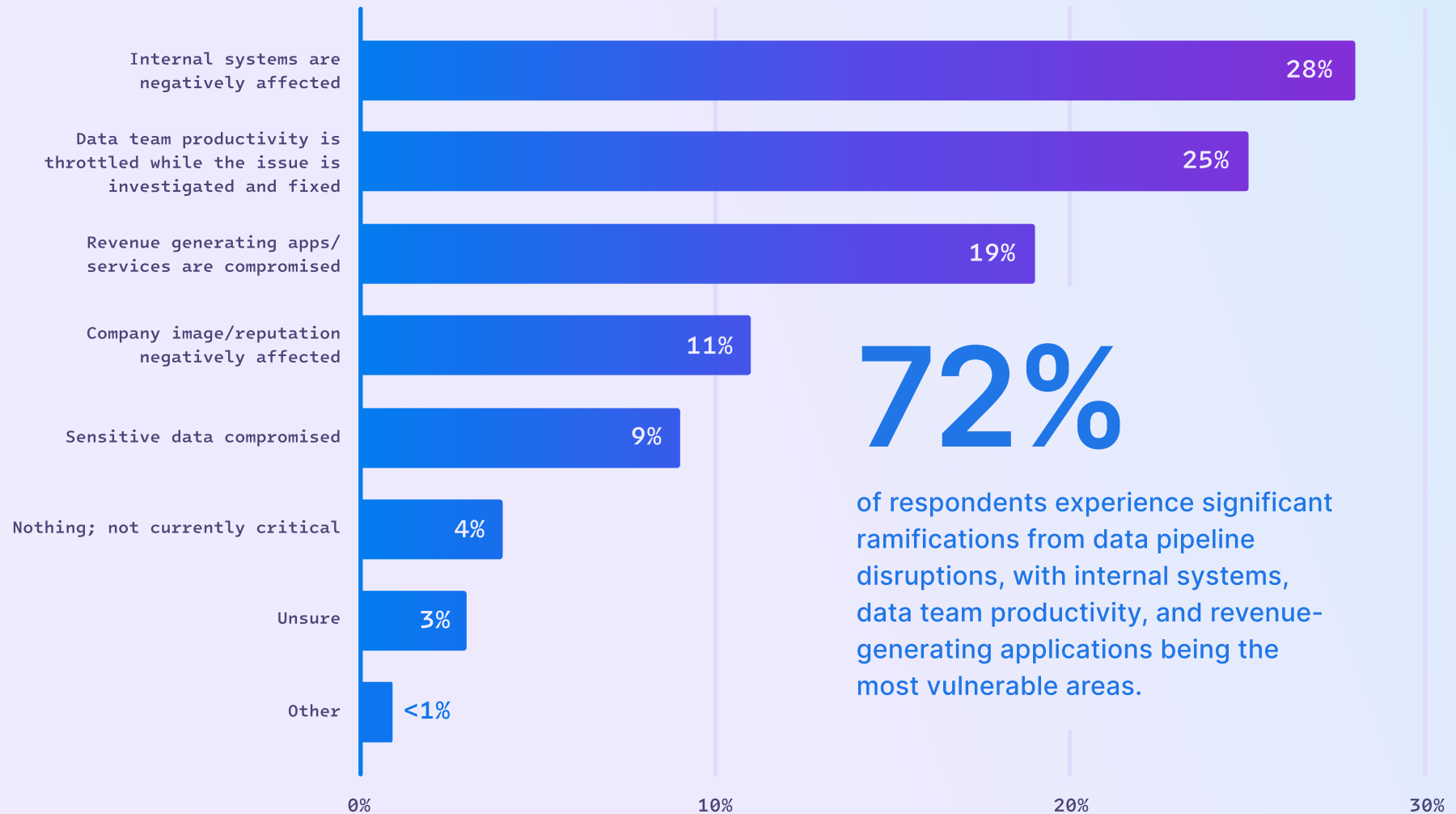
Source: 2023 Apache Airflow Survey, n=801

46%

of Airflow users stress the indispensable role Airflow plays in their business operations, where any downtime can have severe repercussions across the entire organization.

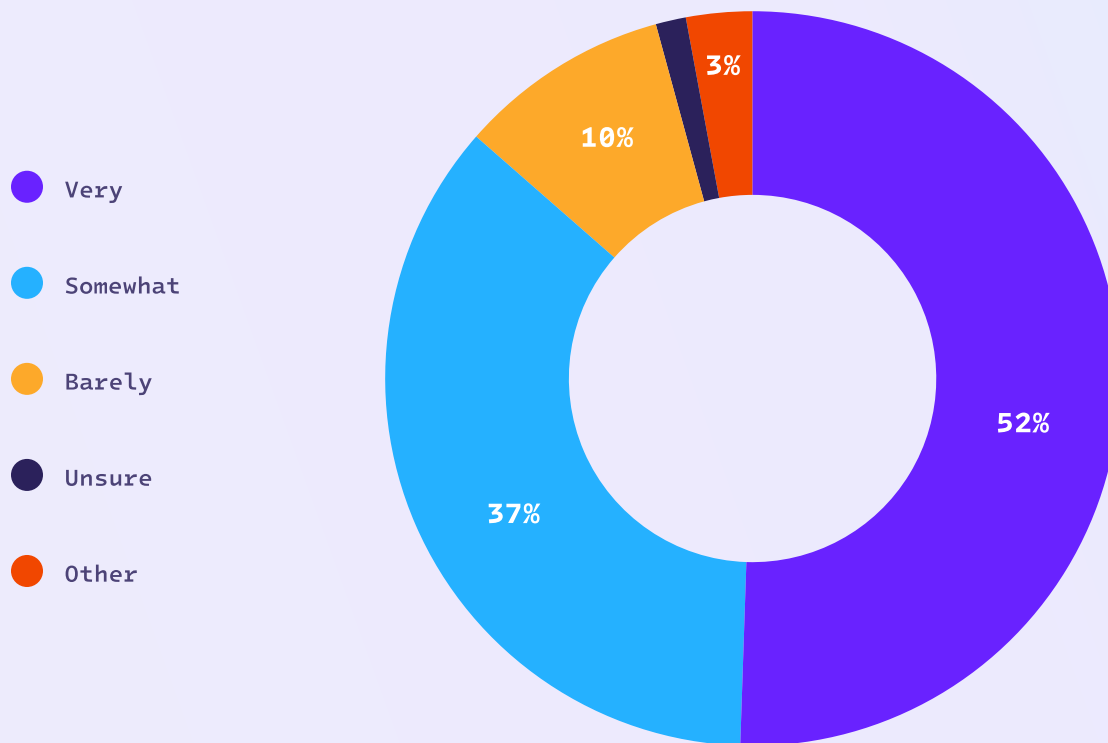
Another **36%** emphasize its vital importance, highlighting how downtime affects numerous teams, underscoring that data pipelines are not merely analytics infrastructure but crucial components of mission-critical business infrastructure.

What happens if your data pipelines go down/data isn't delivered on time?



Source: 2023 Gatepoint Research, n=281

How closely is data delivery/data orchestration tied to your business-critical data?



89%

of respondents acknowledged a strong connection between data delivery/data orchestration and their business-critical data, with **52%** believe it's crucial, underscoring the pivotal role of efficient data management in organizational success.

Source: 2023 Gatepoint Research, n=281

How Teams are Using Airflow

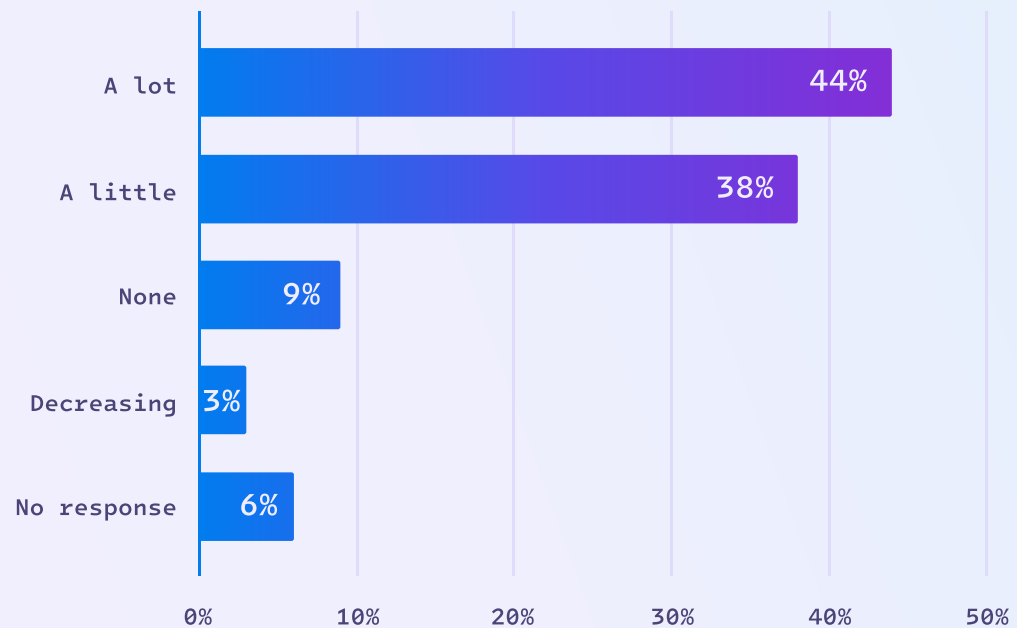
A notable increase in the number of use cases for which teams are using Airflow shows that teams are growing in their confidence of the platform and expanding their vision of what it can accomplish.



82%

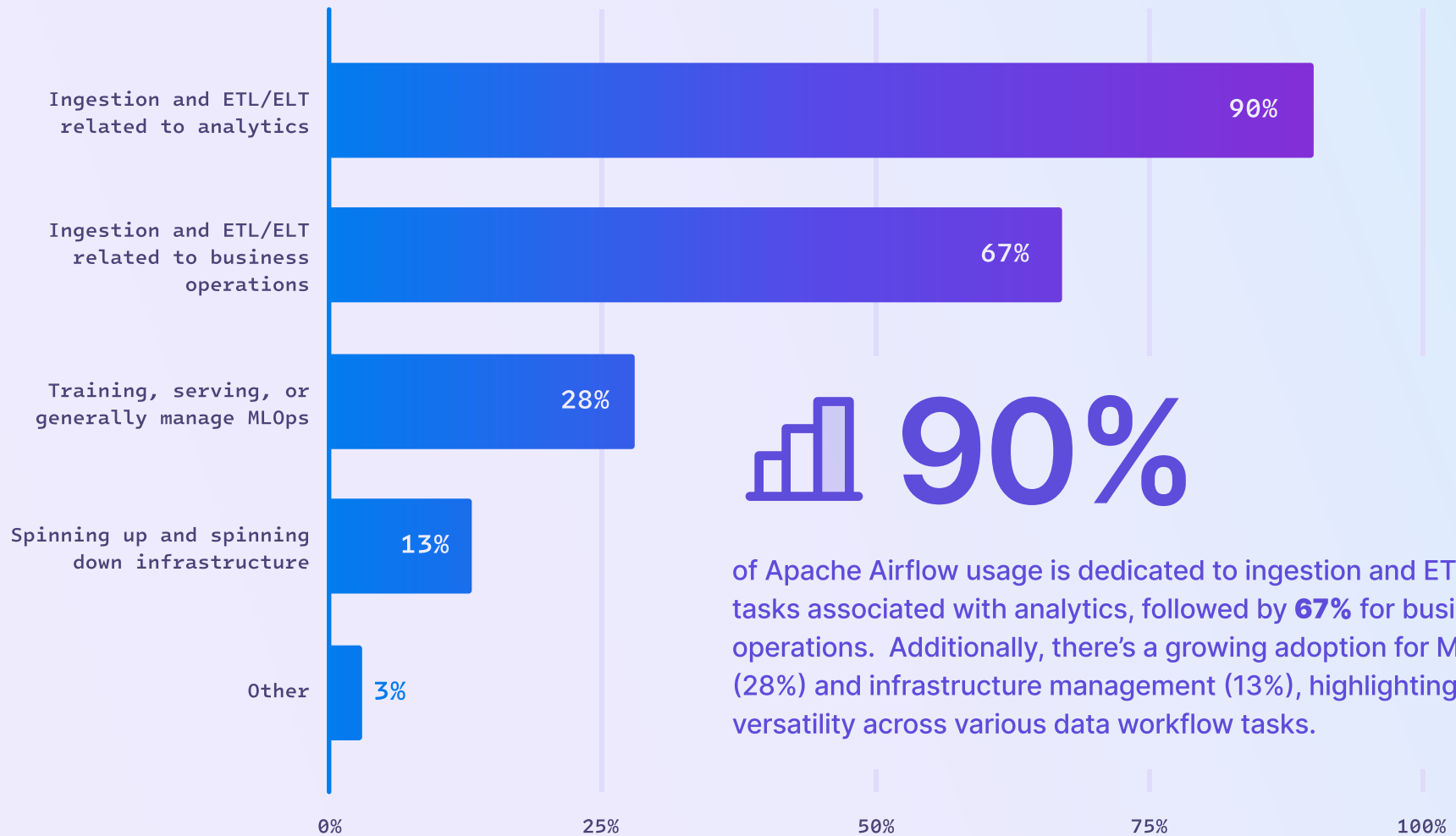
of respondents say their team's Airflow use cases have increased over time, with the majority of those saying use cases have increased substantially.

How have your team's use cases with Airflow grown over time?



Source: 2023 Apache Airflow Survey, n=761

What use cases do you use Airflow for?



Source: 2023 Apache Airflow Survey, n=801



Trend 3: Airflow is Becoming the Standard for ML/AI Pipelines

It's widely recognized that AI and ML are experiencing rapid growth and are becoming key priorities across various industries, with this trend expected to continue into the next decade.

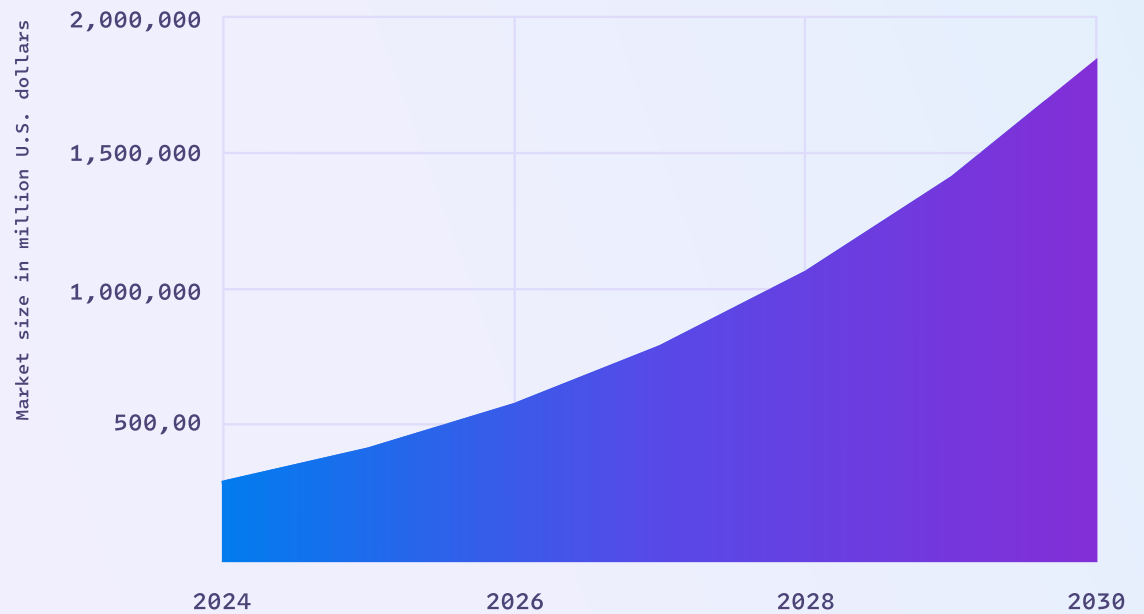
However, particularly noteworthy is the massive increase observed last year in the utilization of Airflow for facilitating AI and ML development. This suggests a growing preference for Airflow over other platforms in powering AI pipelines. Airflow addresses challenges such as isolated development and deployment, lack of coordination and standardization, complex technology landscapes, and compliance issues that often hinder AI and ML innovation.

\$2 trillion

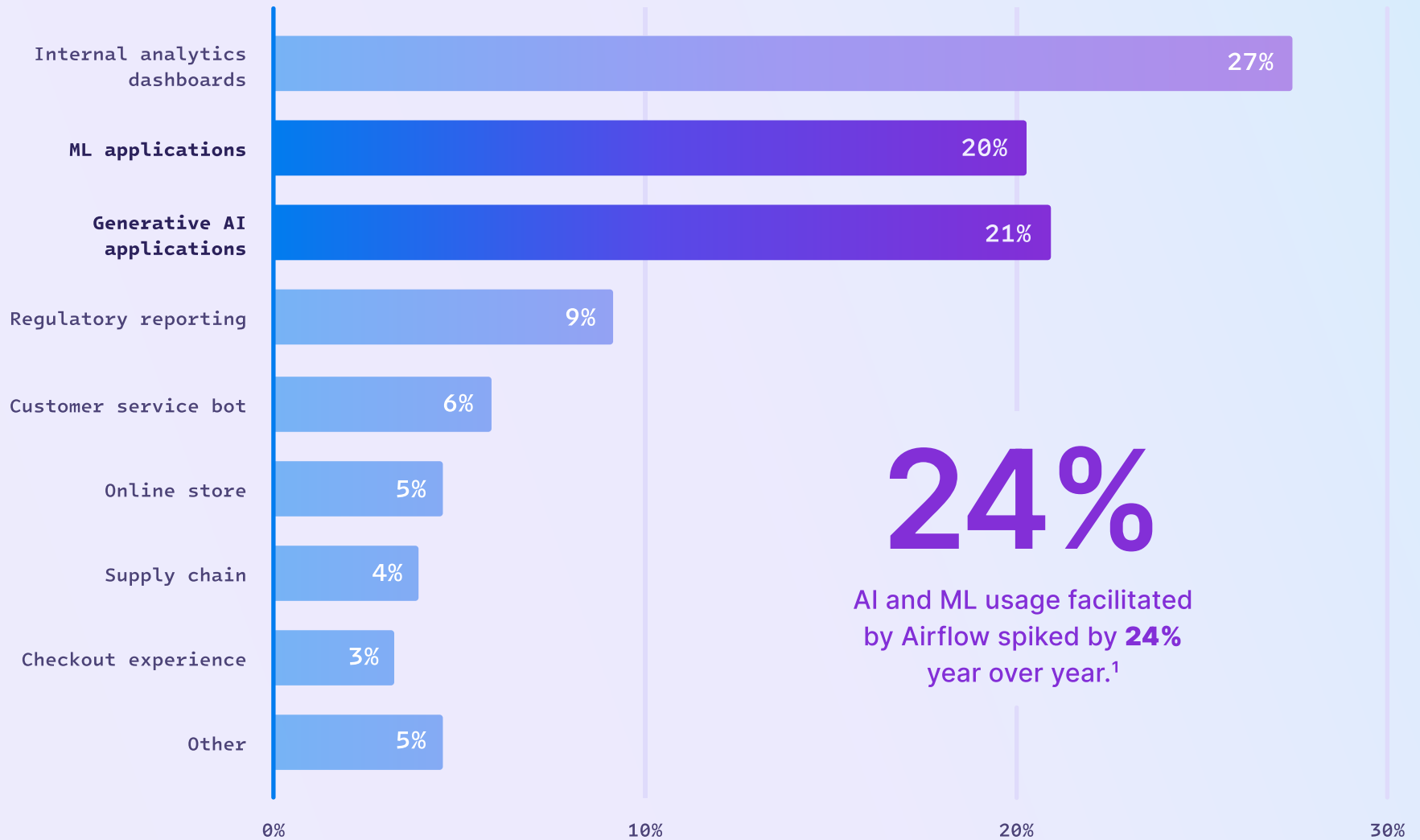
The expected value of the market for AI by 2030

As the global market size for AI is projected to grow by **619%** between 2024 and 2030, Airflow is positioned to be the platform of choice for driving this growth, with ML and AI applications already ranked among the top initiatives companies drive in Airflow.

Forecasted Global Artificial Intelligence Market Size Growth From 2024-2030³



What does, or could, Airflow drive for your organization?



Source: 2023 Gatepoint Research, n=281

How Astronomer Counters Challenges to AI Innovation

As businesses strive to leverage the projected growth of AI, they encounter key challenges such as siloed tools, limited collaboration, fragmented tool stacks, and increasing compliance and operational requirements.

However, an increasing number of businesses are addressing these obstacles by adopting Astronomer—a unified data platform built on Apache Airflow. Astronomer empowers them to:

- Unify and standardize development practices for production-ready AI
- Support next-generation applications with unmatched compute power
- Ensure AI trustworthiness with data lineage
- Accelerate AI development with seamless integrations

Astronomer's Contribution to Apache Airflow

Astronomer has been instrumental in making it easier for individuals and teams to adopt, improve, and take full advantage of the Airflow platform.

Airflow Community

Astronomer has become a force in promoting the use and improvement of Airflow in the community—a space filled with end users and core contributors, designed for data practitioners to learn, grow, and connect. With Airflow Medium, over 33,000 members on Slack, more than 2500 contributors, and meetups happening all over the world, the Airflow community is second to none.

Astronomer is responsible for:



52%

of Commits



50%

of Source Code



8

PMC Members



16

of the Top 25 Committers



31,000+

Users Trained on Airflow

About Astronomer

Astronomer is a unified data platform built on Apache Airflow, designed to unlock the full power of Airflow. While using Airflow, many businesses struggle to find the scalability and availability needed to keep critical applications running, bring data together from the ever-increasing tools in the modern data stack, maintain governance, collaborate between teams, keep development cycles short, and provide required levels of data security and protection.

With Astronomer, organizations seamlessly navigate the intricacies of data operations, leveraging a suite of robust features designed to optimize performance, streamline collaboration, and fortify data integrity →

Maximizing Data Operations: The Astronomer Advantage



Keeping Critical Applications Running

Astronomer allows businesses to meet any data demands with elastic scaling of deployments and the clusters they're hosted on, as well as high availability, minimizing downtime and business disruption with queue-based task execution with dedicated machine types for greater reliability and efficient use of your compute resources.



Unifying Data Across Clouds, Teams and Deployments

Astronomer centralizes DAG, metadata, and issue management, simplifying operations and facilitating effective monitoring. Its built-in data-centric alerting system offers customizable alerts, delivering timely notifications via Slack or PagerDuty for proactive response. With 1,500+ integrations, it seamlessly incorporates all tools within the data stack.



Accelerating Development

Astronomer's tools streamline pipeline deployment, boosting productivity. The Cloud IDE enables swift creation of Airflow pipelines in SQL or Python, ensuring governance and collaboration. The platform provides a seamless local development environment for quick setup, testing, and execution of Airflow. Branch-based deployments facilitate isolated testing for DAGs, while smooth code promotion ensures a secure SDLC.



Ensuring Data Governance

Astronomer is the only managed Airflow service suitable for customers requiring high levels of data security and protection. With enterprise single sign-on, it provides a unified permission model across Airflow deployments, integrated with your enterprise identity provider, as well as role-based access control across workspaces, teams, and deployments. Environment isolation ensures that data is always fully isolated.



438%

return on investment with
payback in < 6 months

55%

increase in compute
efficiency

90%

decrease in task
scheduling delays

20%

reduction time to develop, test,
and deploy new pipelines



“Astro allowed us to get more out of Airflow by scaling up our deployments without the need to increase resources spent on managing the infrastructure. We’ve seen a more stable usage of the infrastructure across all times. This helped us achieve better results, not just in meeting our internal SLAs, but also having more consistency around stakeholders’ expectations.”



Damian Grech

Director of Engineering, *FanDuel*

Ready to Accelerate AI Innovation?

Operationalize and scale AI and ML initiatives, unlock the full power of Airflow to deliver production-ready AI, and accelerate workflow development with Astronomer.

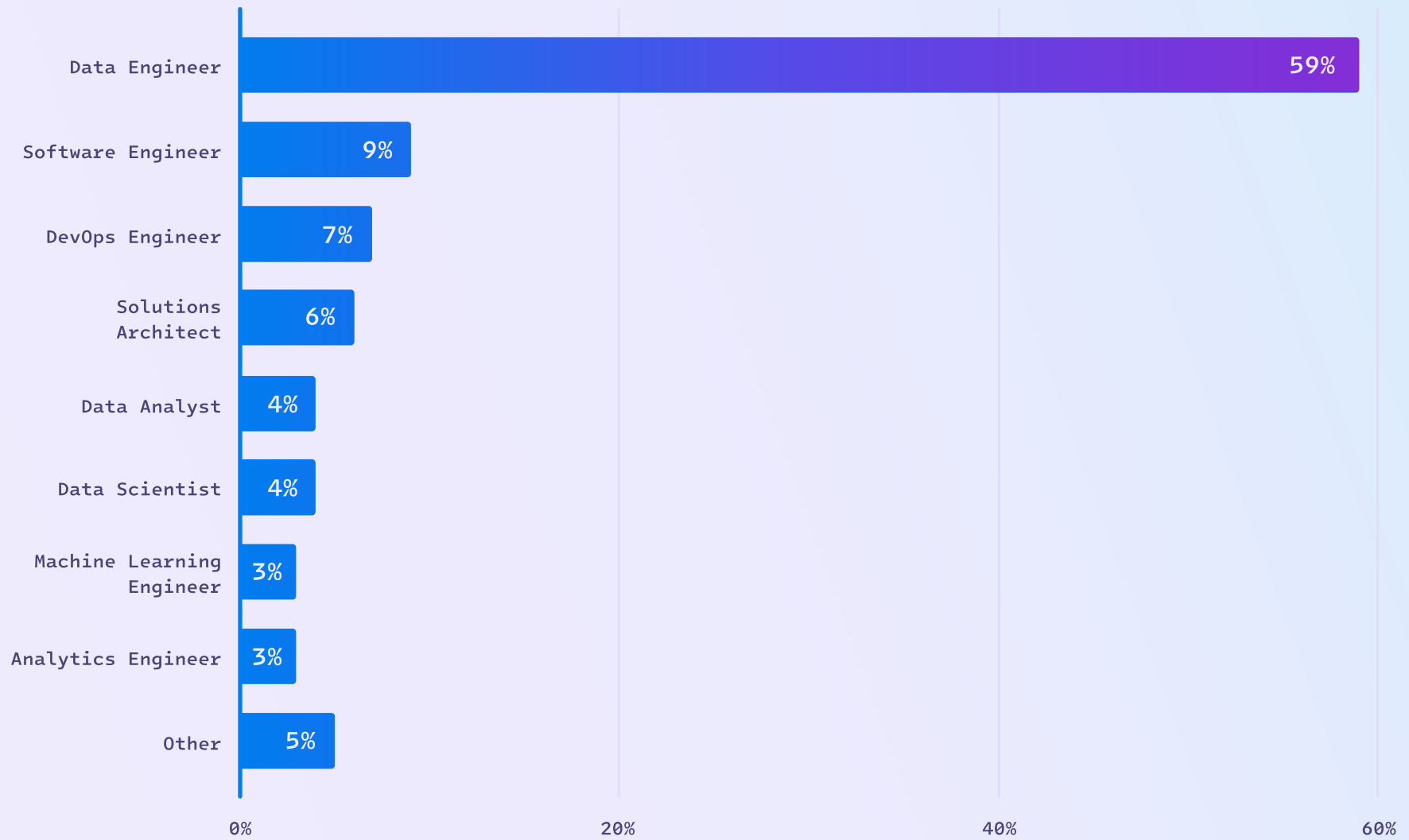
[Get Started →](#)

Methodology

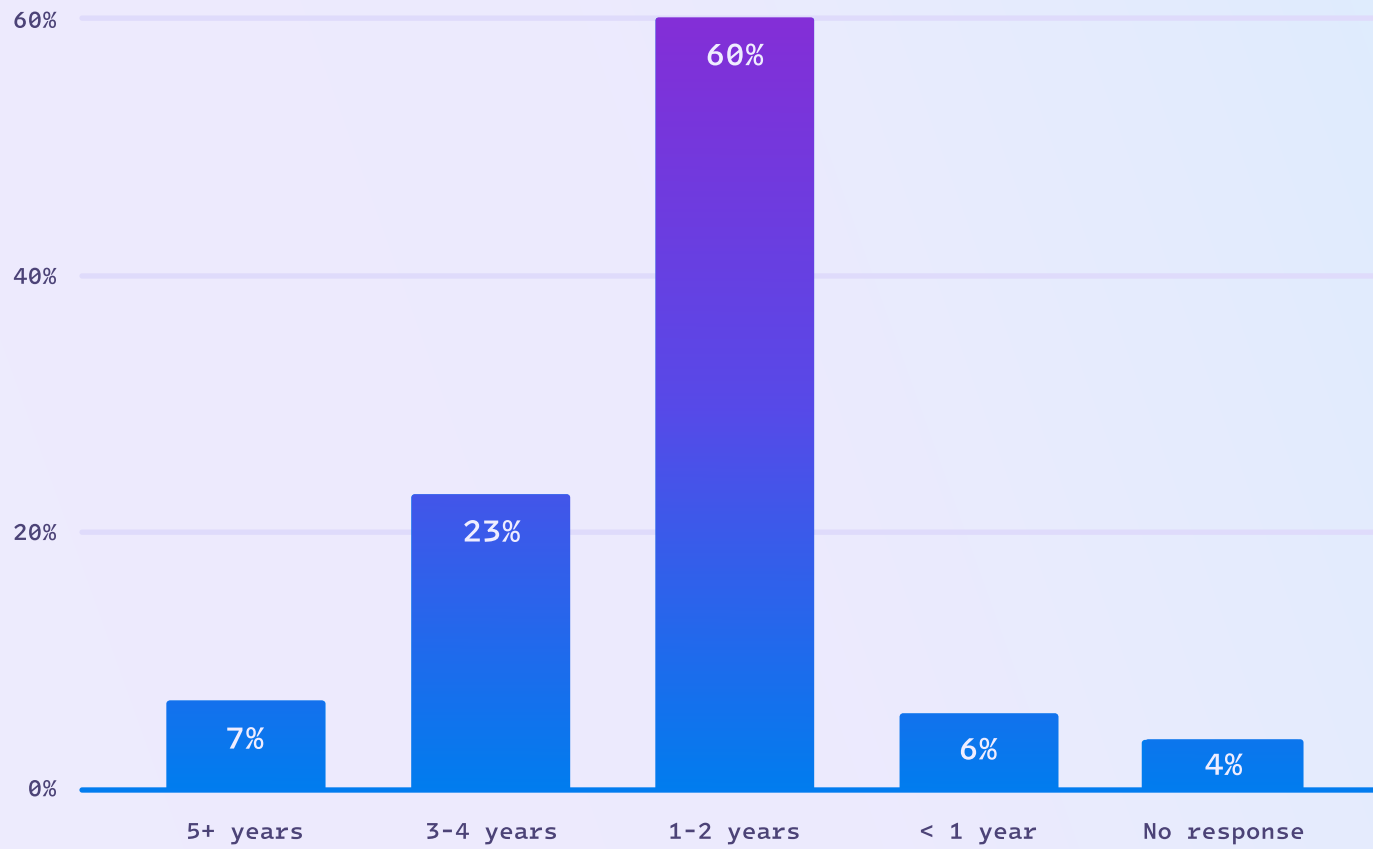
The 2023 Apache Airflow Survey is an annual user survey conducted by the Apache Airflow™ Community, spanning from November 14, 2023, to December 15, 2023. The raw data from the 1,330 captured responses is [available for review](#).

Conducted between June 2023 and January 2024, the Gatepoint Research survey gathered insights from 281 respondents, all of whom were either directly or indirectly responsible for managing their organization's Airflow instance(s). Below, you'll find details on respondents' occupations, company sizes, and levels of experience with Airflow.

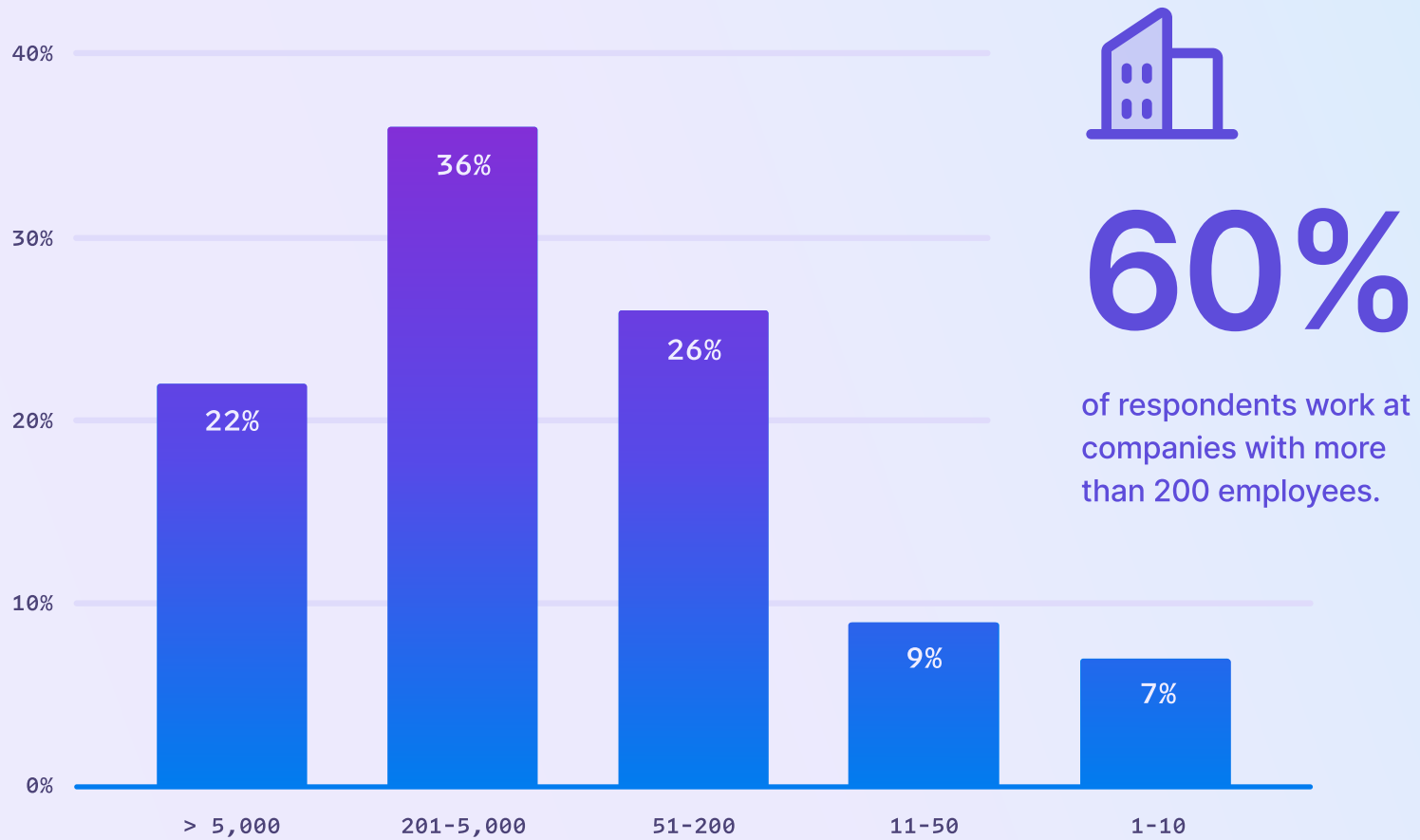
Occupation



Experience with Airflow



Company Size



For any inquiries about the report, [please contact Astronomer.](#)

Additional Sources

1. Astronomer. ["Astronomer Revenue Jumps More Than 200 Percent in H1 2023 as Apache Airflow™ Growth and AI Initiatives Drive Global Demand for Data Orchestration Platform."](#) Astronomer press release, September 7, 2023. On PR Newswire.
2. ["Apache Airflow,"](#) GitHub.com, December 2023
3. ["Artificial Intelligence \(AI\) Market by Component \(Hardware, Software, and Services, Deployment \(Cloud and On-Premise\), Application \(Virtual assistants/Chatbots, Forecasts & Modeling, Text Analytics, Speech Analytics, Computer Vision, Predictive Maintenance, and Others\), and End-User Industry \(BFSI, Government, Aerospace & Defense, Automotive, Healthcare, IT & Telecom, Manufacturing, Education, Retail & E-commerce, Energy & Utilities, Media & Entertainment, and Others\): Global Opportunity Analysis and Industry Forecast, 2022-2030,"](#) Next Move Strategy Consulting, January 2023
4. [Apache Airflow Survey 2023](#)

Copyright © Astronomer 2024. All rights reserved.

Apache, Apache Airflow, Airflow, Apache Kafka, Kafka, Apache Spark, Spark, and associated open source project names are trademarks of the Apache Software Foundation.