state of Air Flow 2025

ASTRONOMER

Table of Contents

Executive Summary	3
Popularity and Relevance	5
Growth	7
Al powered by Airflow	8
Enterprise-Grade Airflow with Astronomer	13
Astronomer and Open-Source Airflow	
A New Era: Airflow 3.0	18
About the Report	21



Executive Summary

Every tech and business leader woke up this morning thinking about data, and how to turn it into a competitive advantage.

Whether delivering customer-facing data products, enriching advanced analytics use cases or getting past the prototype stage and into production with AI, the world's largest enterprises know that unprecedented opportunities exist within their data. But simply collecting data isn't enough. It must be extracted, aggregated, cleaned, transformed, and backfilled to fuel the latest AI models—all within an ever-more complex ecosystem of specialized tools that can quickly turn to chaos without the right orchestrator. The most successful businesses cut through the noise by unifying their DataOps with orchestration at the foundation.

To better understand how to seize the Data Opportunity, this report features insights from the engineers that do the work that goes beyond simple analytics: bringing AI solutions to market and taming massive troves of data to redefine what a data product can be. The 2025 State of Airflow report is built from more than 5,000 respondents, making it the largest data engineering survey to date. Combined with data collected from some of the world's leading companies running Apache Airflow® at scale with Astronomer, we've uncovered the key trends shaping the future of data.



Three core themes have emerged:

of users say Airflow is important to their business *N=5255*

Airflow Is Business-Critical

Airflow isn't just extremely popular with engineers, receiving a >90% recommendation rate (N=4889), it's a core component of the most critical data operations at leading enterprises. More than half of users in organizations with 50,000+ employees report using Airflow for business-critical workloads.

OVER85%

of users expect to use Airflow for more external or revenue generating solutions next year

N=5255

Unprecedented Growth of Revenue-Generating Solutions

Data orchestration is no longer just about internal dashboards or operational efficiency, it is driving direct business impact. Over 85% of surveyed users expect an increase in external-facing or revenue-generating solutions built on Airflow in the next year.

OVER 35%

of users with 5+ years of experience use Airflow for ML/AI

N=481

Orchestration-first organizations deliver production AI while others stagnate

As companies scale, so does their expansion of Airflow use. What begins as a framework for ETL and ELT evolves into a critical platform for deploying machine learning (ML) and generative AI (GenAI) initiatives into production.

Popularity and Relevance

In recent years, Airflow has firmly established itself as the connective tissue of the modern data stack.

The last major update to Airflow was the release of Airflow 2.0 in 2020. In the time period since, we can clearly see a surge in Airflow popularity, usage and community growth by comparing the metrics from November of 2024 and the same time period in 2020. As of November 2024, Airflow boasted more than:

- **31M** monthly downloads (less than 900K in 2020)
- 3K contributors (around 1.3K in 2020)
- 29K pull requests (about 10K in 2020)
- 77K organizations using Airflow (approximately 25K in 2020)

The findings in this report suggest the project will continue to grow in popularity and criticality in the larger technology ecosystem. Over 90% of survey respondents say Airflow is important to their business, with **47.9%** reporting that it powers **business-critical use cases** within their organization. As new AI solutions and data products demand increasingly accurate and extensive data, this number will keep rising.

Engineers in enterprises with 50,000+ employees confirm what we see at Astronomer every day: when operating an **enterprise at scale**, Airflow's role, and its importance, grows alongside the business. It's no surprise that within this group, **53.8%** of users classify Airflow as a **businesscritical technology**.



Airflow's dominance as a foundational orchestration platform isn't just reshaping businesses, it's shaping careers. Data teams recognize its value, with **90%** of survey respondents stating that their Airflow expertise has already had or will have a **positive impact on their professional growth** in the next 5 years.

90%

of users say knowing Airflow has or will have a positive impact on their careers

I=5255

Do you think having a good foundation in Airflow and DAG writing will have a positive impact on your career in the next 5 years? **35.2%** Yes, Airflow is central to my job and career

25%

Yes, I am currently pursuing or plan to pursue more education in Airflow for the purpose of career growth

29.8%

Yes, knowing Airflow already has had a positive impact on my career

ASTRONOMER

Growth

It's not just the criticality of use cases that grows with enterprise size — as companies scale, they bring larger data volumes, more platforms in their data stack, and increasingly complex use cases.

Their investment in Airflow expands as well. In fact, large enterprises are pushing the limits of hyper-scale Airflow infrastructure: **over** - -**20% of companies with 50,000+ employees** report running **20 or more production Airflow instances**, compared to **13.0%** of mid-sized businesses and just **5.4%** of small businesses.

What industry leaders and fast-growing startups have in common is a shift beyond traditional Airflow use cases, moving from powering internal BI dashboards to also orchestrating **external**



and revenue-generating solutions (for example: Advanced Driver Assistance Systems processing more than a petabyte of data a week, continuously deploying new features to users of the largest business social network, or meeting strict auditing and compliance standards in the lightning-fast crypto space). This momentum is only accelerating: over 85% of Airflow users expect these business-driving applications to grow in the next year, illustrating a massive opportunity for enterprises that have already fully embraced Airflow throughout the business.



At Astronomer, we have a front-row seat to the incredible growth of enterprises that have partnered with us. Astro customers saw their monthly **compute** hours **grow by an average of 97.3%** from 2023 to 2024, while **more than doubling their average monthly successful task count**, a testament to the increasing scale and impact of Airflow-driven operations. When an enterprise invests in a strong data orchestration foundation driven by Airflow, there is compelling evidence that it enables them to expand their Airflow operations into new use cases throughout the organization.

+97.3% Avg. monthly compute hours of Astronomer customers, 2023 vs 2024

+103.7% Avg. monthly successful task count of Astronomer customers, 2023 vs 2024

AI powered by Airflow

While traditional use cases like data ingestion, and ETL/ELT for analytics and business operations are still very popular with Airflow, we're seeing a rapid expansion into **MLOps** and **GenAI**.

The survey data confirms this shift: **21.4% of users already use Airflow for MLOps, while 8.2% have adopted it for GenAl use cases**. These numbers are even higher for experienced users, and expectations for adoption within enterprises are accelerating.

What sort of use cases do you use Airflow for?

This evolution isn't surprising. **Core data engineering**, including ETL and ELT, provides the data that forms the **foundation of advanced Al applications**. Before building Al-driven solutions, companies must first collect, clean, and prepare their training and fine-tuning data.



Users with 5+ years experience N=481, multiple responses allowed

Airflow 3.0 – which will be released in April of 2025 – will introduce even **more data-awareness features** tailored to these needs, reinforcing Airflow's role as the backbone of modern AI pipelines.



Airflow Maturity Drives AI Adoption

The longer engineers use Airflow, the more they use it for Al. Among users with 5+ years of experience, **over 30% run MLOps workloads, and 13.3% use Airflow for GenAl.** This reflects a clear maturity trajectory, as teams gain expertise, Airflow becomes the go-to orchestration layer for Al-driven innovation.



Airflow Future-Proofs your AI Investments

Al will continue to evolve. There hardly is a day without a new groundbreaking model release, and this will put new demands on data infrastructure. With Airflow as the flexible, extensible core of the modern GenAl stack, enterprises aren't just keeping up, they're future-proofing their ability to innovate, no matter what comes next.



Astro Customers Lead the AI Wave

Organizations that use Astro by Astronomer are ahead of the curve: 55% of Astro customers already use Airflow for ML and Al. As customers mature their operations on Astro, their Al initiatives tend to expand, with **69% of those who have been our partners for two or more years** using Airflow for Al and ML. To further accelerate this use case adoption, we've recently launched the first-ever Airflow course for GenAl and created a set of reference architectures designed to help enterprises orchestrate Al pipelines with confidence.

Data orchestration is pivotal for organizations aiming to implement productionlevel AI solutions, and Airflow is the most broadly used orchestration software with a proven track record of supporting AI and MLOps. By systematically managing and integrating data from diverse sources, orchestration ensures that AI models receive accurate and timely information, which is essential for effective training and decision-making. Airflow has become integral in overseeing these complex workflows at leading enterprises, enabling the seamless ingestion, processing, and transformation of data. This structured approach not only enhances the reliability and scalability of AI applications but also accelerates the deployment of AI-driven innovations across various industries. Put simply-the linchpin to getting AI solutions right is an understanding of data orchestration fundamentals. That's why it's not surprising to see AI and MLOps workloads increase among veteran Airflow users (5+ years) and drastically increase among Astro customers, especially those who have been partnering with us for more than 2 years, benefiting from access to the Airflow expertise and best practices guidance at Astronomer.

Let's break out three examples of what production AI built on a foundation of Airflow orchestration can look like.



One of the largest global, online travel agencies providing reservation services for millions of properties in more than 200 countries, hotels, as well as flights and activities.

The Company provides travelers with a seamless experience by integrating all aspects of a trip into a single platform. To do so, The Company processes petabytes of data daily, generating 100 terabytes of machine learning predictions each day. Airflow, managed by Astro, serves as the backbone for orchestrating critical workloads, from analytics and business intelligence to machine learning and Al.

With Astro solidified as a stable platform for running Airflow pipelines, the team is able to extend the use of orchestration into new use cases, especially in applications of machine learning and Generative AI. The search ranking system, front and center on their website, relies heavily on Airflow to provide the flight and hotel listings that are essential for every traveler. One recently launched AI solution helps travelers brainstorm their next trip with a conversational AI planner. Whether in the ideation phase of where to even visit, or locking down the exact desired specifications of a hotel, the AI solution leverages Astro to support every stage of vacation planning and is an exciting new way for travelers to build out their itinerary.

By entrusting Astro with the management of its data workflows, The Company has been able to focus on innovation, driving business growth while processing billions of transactions and predictions daily with ease.



An international leader in on-demand, custom printing leverages a robust and ever-changing network of suppliers.

To keep up with growing demand and maintain a competitive advantage, The Company needed to streamline its product onboarding process across multiple languages and categories, a traditionally manual, time-intensive task.

The Company's data team faced a core challenge: onboarding thousands of new products from suppliers across Europe and the United States. Each product required descriptions, translations, and images that aligned with The Company's brand tone. The existing process was slow and manual, limiting growth.

The data team had previously adopted Astro for handling more traditional use cases like data warehousing and workflows for business operations. With those fundamentals in place, they decided that Astro was the obvious choice to orchestrate the GenAI-powered content generation process, leveraging Airflow's capabilities to integrate third-party APIs, manage batch processing, and monitor data quality. For example:

- **1. Supplier Data Integration:** Airflow pipelines ingest and standardize supplier product data.
- 2. GenAl Content Generation: Airflow then orchestrates calls to OpenAl's GPT-4 to generate product descriptions in English, maintaining The Company's brand voice.
- **3.** Automated Translation: Airflow tasks then trigger multilingual translations into Dutch, French, Spanish, and other languages.
- **4. Human-in-the-Loop Review:** Content managers review Al-generated descriptions via a Google Sheets interface that acts as a simple front-end to Airflow DAGs.
- 5. Automated Publishing: Upon approval, Airflow pipelines publish the content to The Company's CMS.

The choice of technology for production GenAl was obvious from the beginning: Airflow provides the batch-processing capabilities, running on Astro to ensure reliability, scalability, and streamlined operations. The results were operationalized Al at scale:



10x Faster Product Onboarding

Thousands of new SKUs were launched in months instead of years.



Simplified User Interaction

A Google Sheets-based interface allows non-technical users to trigger complex Airflow pipelines with a single click.



Seamless AI Model Integration

Switching between OpenAl and Gemini models is straightforward, ensuring costeffective AI operations.



A bleeding-edge, US-based financial services firm that allows accredited investors from across the globe to invest in pre-IPO companies.

As part of their commitment to providing clients with best-in-class service they are regularly launching new apps and services. Because The Company standardized on Astro for data orchestration, they are able to leverage AI to significantly increase their efficiency in processing support requests.

Previously, their support team had to manually review and process every ticket, filling out forms and responding without AI-driven assistance. But with Astro reliably delivering clean, structured data to Snowflake, The Company was able to rapidly build AI-powered

workflows that analyze support tickets and generate response suggestions significantly reducing manual effort and accelerating resolution times.

Astro's ease of building and managing pipelines ensured the organization could quickly deploy this Al solution. Additionally,



to maintain compliance and prevent PII exposure, they leveraged Astro to extract, clean, and securely return data to Snowflake before feeding it into their AI models.

By standardizing on Astro, The Company was able to accelerate AI adoption, reduce operational overhead, and improve customer support efficiency—all while ensuring data security. Relating to the picture above, The Company was able to create the ticket processing app in one week with less than 600 lines of additional code because Astro already connected to all their systems and they had standardized on Astro as the way to move data between different applications in their data stack.

Enterprise-Grade Airflow with Astronomer

21.4% of Airflow users identify infrastructure limitations as a major obstacle to delivering business value.

Astro addresses this challenge by providing scalable infrastructure, which, according to a Forrester report, reduces infrastructure management workload by **75%** for customers.



Multiple responses allowed

Looking at how enterprises, in particular the Fortune 500, leverage Airflow in the context of their whole data ecosystem, another trend emerges. Facing more complexity than ever in moving their data into, through and out of a multitude of tools, they heavily rely on Python-, Bash- and KubernetesPodOperators, as well as entirely custom operators to use Airflow for their data operations, despite many vendors advertising "out-of-the-box" integrations between popular tools.

In fact, **64.4%** of Airflow users in large enterprises report using **Bash- and PythonOperators**, and **16.2%** of them leverage the **KubernetesPodOperator**, alongside **30.4%** writing **custom operators and hooks**. What these methods of defining Airflow tasks have in common is that they allow engineers to write custom code to manipulate data and interact with any tool that has an API in a fully customizable way. This data underscores the importance of the fully flexible orchestration layer that Airflow provides, giving its users an edge in adapting What other

tools in the

data stack do

their data strategy to a quickly evolving market. Among Astronomer customers the flexibility and agility is even more pronounced with **95.3**% using the PythonOperator, **22.3%** the **KubernetesPodOperator**, and **51.4%** using the **BashOperator** to run custom code. Put shortly: Airflow gives enterprises the freedom to operationalize their data however they choose, without ever being dependent on any one vendor.



How do you usually interface with other services from your Airflow DAGs? (50k+ employees)

Adding to this complexity, the data ecosystem remains fiercely competitive: Snowflake (\$SNOW), Databricks (DBX), and BigQuery are going head-to-head for dominance. This trend is evident both in the broader Airflow community, where \$SNOW (28.0%), DBX (29.0%), and BigQuery (27.6%) have nearly equal user shares, and among our internal customers, where Snowflake currently leads in overall adoption, while BigQuery has the highest total task count and total compute time. Market analysts may be looking for a clean winner in the data platform wars, but enterprises have clearly chosen a strategy of **data platform diversity** for their analytics and AI – just like enterprise purchasing preferences previously determined that **multi-cloud strategies** would far outpace standardization on any single cloud provider.

As **heterogeneous tech stacks** increasingly become the standard for enterprises, Airflow becomes a linchpin in achieving strategic leverage by providing flexibility via cost-effective data movement.

	\$SNOW	DBX	BigQuery
Customer Adoption	49.1% (+15.5% y/y)	21.8% (+2.7% y/y)	26.5% (+6.4% y/y)
Successful Task Count	++	+	+++
Total Compute Time	++	+	+++

Internal customer data

Leading enterprises refuse to lock into a single vendor, opting for multi-cloud and diverse data platform approaches to stay agile and take advantage of the latest innovations. Airflow's tool-agnostic nature makes it the perfect orchestration layer for this dynamic and rapidly changing environment.

With data and processes fragmented across an ever-increasing number of tools, leading enterprises are shifting toward unified DataOps strategies to regain control. And no layer in the modern data stack is better suited as the foundation of DataOps than the orchestrator, with its deep insight into every tool in the data stack, as well as the data and metadata being processed.

To put it simply: **DataOps has to be built on orchestration—and orchestration is synonymous with Airflow.**



As further evidence of Astronomer's role in the modern data stack, a relatively young package for using Airflow with the transformation technology dbt Core, exploded in popularity in 2024. Released by Astronomer and free to use by the open-source community, Astronomer Cosmos greatly enhances the experience of using dbt Core for data transformation in Airflow pipelines. The download numbers speak for themselves: Cosmos is currently being downloaded 13 million times per month, making it the most popular way to run Airflow and dbt Core together.

Astronomer and Open-Source Airflow

One of the biggest challenges teams face in delivering value with Airflow is a **lack of people resources**, reported by 34.2% of survey respondents, followed by a **lack of Airflow literacy** in other teams, which impacts 25.4% of users.

Where do you get information about Airflow releases, features, & changes?

To help tackle these challenges, Astronomer continuously publishes educational resources for Airflow users at all levels. **22.3%** of all survey respondents, including open-source users, said they rely on **Astronomer's documentation** to learn about new Airflow features.

Asked what they would like to see more of from the Airflow community, Airflow users affirm our dedication to expand on successful Astronomer resources.

- 50% would like to see clear documentation and examples: Our guides and eBooks like the Ultimate Guide to DAGs are designed to elaborate on Airflow features while providing plenty of code examples. The Astronomer Registry is a comprehensive resource documenting Airflow providers and their modules.
- 46.9% of users are asking for best practice guidance: The reference architecture section in our documentation as well as dedicated best practice webinars are Astronomer resources offering opinionated guidance in how to st



Multiple responses allowed

resources offering opinionated guidance in how to structure Airflow projects for different use cases.

 42% would like to attend more virtual events: Astronomer runs regular webinars covering everything from new features to deep dives into how to use Airflow with different data tools.

- 34.8% are asking for getting started resources: Whether you're new to Airflow or need a refresher, we offer resources like the Airflow 101 eBook and Get Started documentation to get you up and running. Additionally, the Astronomer Academy offers full-length Airflow courses and certifications.
- 34.4% of users are looking to connect at local meetups and events: We support and run regular Airflow meetups and other events in many major cities.



What would you like to see from the Airflow community?



A new era: Airflow 3.0

Astronomer is the driving force behind Apache Airflow and Airflow 3.0 is going to take the project to new heights.

In our role we invest heavily in the future of the project to ensure it remains the best tool for data orchestration. We work closely with the Airflow community to make data-driven development decisions, with fast feedback cycles from Astronomer customers and our team of experts. As you'll see, insights gathered from the Airflow survey directly inform the direction of new features and improvements.

We are especially excited for the upcoming release of Airflow 3.0 this Spring (now with Dark Mode), delivering multiple long-awaited features around the 3 core themes:



Ease of Use

- **DAG Versioning:** One of the most requested features (over 50% in the 2023 Airflow survey, N=747). Airflow now fully retains previous DAG versions and their run history, enhancing observability and developer workflows.
- **UI Modernization:** 39.0% of users are asking for updates to the Airflow UI. Version 3.0 will introduce a more modern and intuitive experience, built on the React framework.



Run Anywhere in Any Language

- Airflow 3 lays the foundation for executing tasks on any machine, on-prem or in the cloud, triggered by events across the data ecosystem. It also introduces a proof of concept for defining tasks in languages beyond Python, greatly improving data team agility and facilitating migration from legacy systems to Airflow.
- A new task execution interface significantly strengthens Airflow's security posture and lays the foundation for executing tasks on any machine, on-prem or in the cloud, triggered by events across the data ecosystem.
- Airflow 3 also introduces a proof of concept for defining tasks in languages beyond Python, greatly improving data team agility and facilitating migration from legacy systems to Airflow.



Expanded Data Awareness

- Data awareness features, introduced in Airflow 2.4 and greatly expanded in Airflow 2.9, already enjoy the highest adoption rate of all features at 48.0% of all users.
- Airflow 3 will introduce features that include a whole new way to build workflows, enabling intuitive, data-driven pipelines. These advancements align with Astronomer's plans to enhance data observability in Astro.
- Run anytime: Airflow 3.0 will introduce event-driven orchestration, which enables workflows to kick off the moment something important happens-no more waiting around for a scheduled run. That means faster insights, less wasted compute, and the ability to power real-time mini-batch use cases like fraud detection or supply chain updates. This feature is being built right on top of Data Assets, making event-driven workflows seamless while keeping all the flexibility Airflow is known for.

Of course, Airflow will continue to evolve, and the 2024 Airflow Survey gathered insights on the next features practitioners want to see to enhance their productivity and experience. One standout request, cited by a full **third of users**, is **data cataloguing**. At Astronomer, we're thrilled to see this strong interest in observability features, as it reinforces our direction in building a comprehensive observability suite **Astro Observe** which has been generally available since February 13.



N=5255, Multiple responses allowed

Airflow 3.0 represents a major step forward, cementing Airflow's role as the standard for orchestration, driven by the needs of its community and Astronomer. Astronomer offers the next-level Airflow experience, delivering on observability and building out the DataOps platform for growing enterprises. We can't wait to see the data products our customers will build running Airflow 3.0 on Astro!

About This Report

This report combines insights from the 2024 Apache Airflow[®] survey with data from Astronomer customers running Airflow on Astro at scale, providing a comprehensive look at how Airflow is used in production today.

AIRFLOW SURVEY 2024 RESPONDENTS

Between October 21 and December 23, 2024, Airflow users worldwide participated in a 50-question survey covering how they run Airflow, develop pipelines, and asking them for feedback on existing and future Airflow features. You can find the full survey results here.

A huge thank you to the 5,255 Airflow users who responded, quintupling last year's engagement and making this the largest Airflow survey to date!

We also want to extend our gratitude to everyone in the Airflow Community for their contributions to this thriving open-source project.



Airflow's reach continues to expand across industries and experience levels:

- 9.2% of respondents are seasoned Airflow users with 5+ years of experience, showing a healthy, high-expertise user base
- 38.1% have one year or less of experience, highlighting a rapid influx of new users
- The majority (61.5%) hold the title Data Engineer and the most common level is that of an individual contributor (68.0%), underscoring how this survey gets insights directly from practitioners using Airflow at work every day.
- A growing number of other professionals are leveraging Airflow, including **Analytics Engineers** (6.4%), **Solution Architects** (5.3%), and **DevOps Engineers** (5.1%), which reflects expanding use cases, Airflow is not just a tool for ETL and ELT anymore, it orchestrates all your pipelines, including an ever growing number of AI applications.



ASTRONOMER DATA

At Astronomer we work closely with customers to make their Airflow experience even better and collect anonymized telemetry on how our customers are using Airflow. Combined with our partnership with select customers and data professionals tackling challenges in the largest companies on the planet we gained unique insights into how Airflow is used at extreme scales.

This real-world system usage data is collected continuously at Astronomer and served into internal dashboards, all using Airflow pipelines, running on Astro!



The unified DataOps platform for modern data teams

Astro, powered by Apache Airflow[®], is built on the principle of consolidating fragmented layers, such as observability, ML and Al Ops, into an orchestration-first, easy-to-use, enterprise-ready solution. This approach allows customers to benefit from purpose-built additional tooling that encourages software best practices, increases pipeline reliability and security, and accelerates development cycles.

Try Astro for Free

Book a Demo

ASTRONOMER