

Demystifying Apache Airflow – Separating facts from fiction

Shubham Mehta

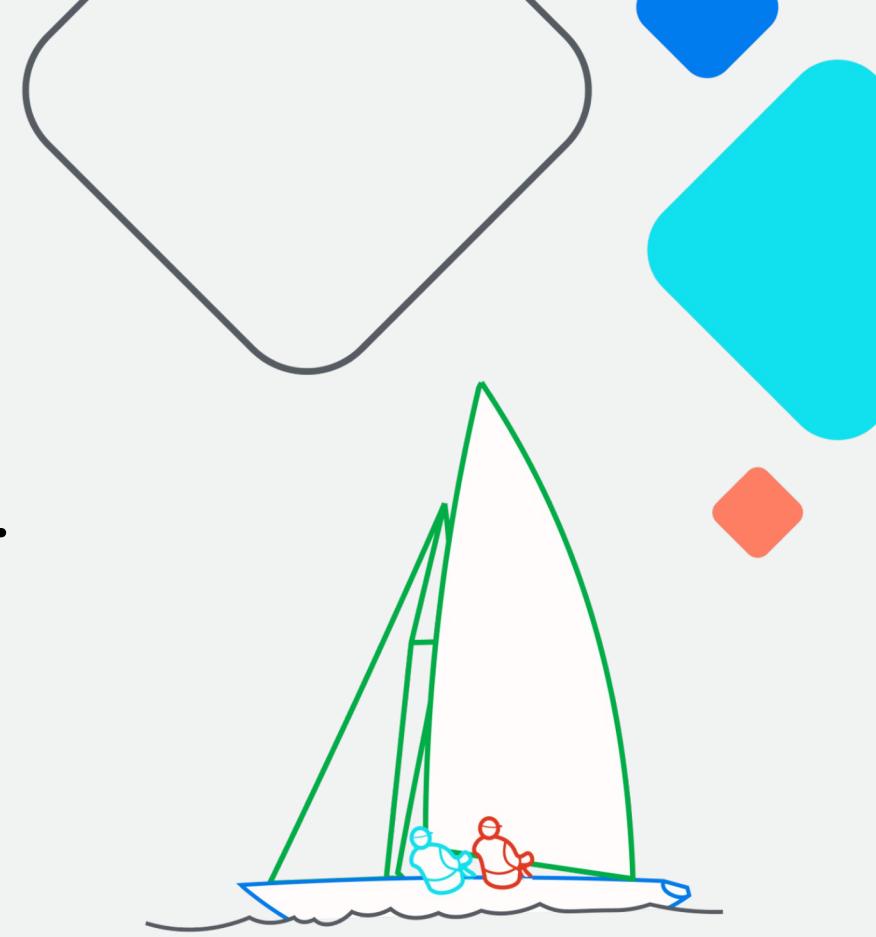
Sr. Product Manager

Amazon MWAA

Uma Ramadoss

Sr. Specialist Solutions Architect

Amazon Web Services



 **Airflow Summit**

Let's flow together

September 19-21, 2023,
Toronto, Canada



Agenda

Top critical feedbacks

Analysis of the challenge

Opportunities

Getting started with Airflow

Critical Feedback 1 - Airflow is not enterprise ready



Why is it a myth?



Security

- Airflow security model
- Dedicated security team
- 25 CVEs resolved in 12 months

Airflow security policy



<https://github.com/apache/airflow/security/policy>

Airflow security Model



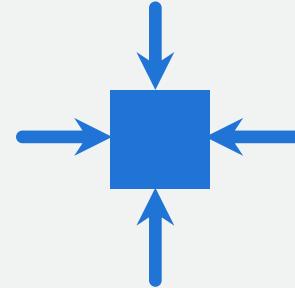
https://airflow.apache.org/docs/apache-airflow/stable/security/security_model.html

Why is it a myth?



Security

- Airflow security model
- Dedicated security team
- 25 CVEs resolved in 12 months



Extensibility

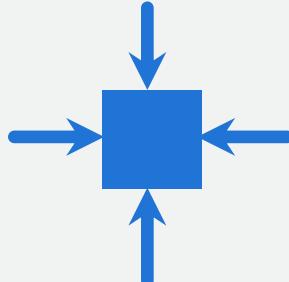
- Airflow V2 scheduler
- Provider packages
- Public interfaces

Why is it a myth?



Security

- Airflow security model
- Dedicated security team
- 25 CVEs resolved in 12 months



Extensibility

- Airflow V2 scheduler
- Provider packages
- Public interfaces

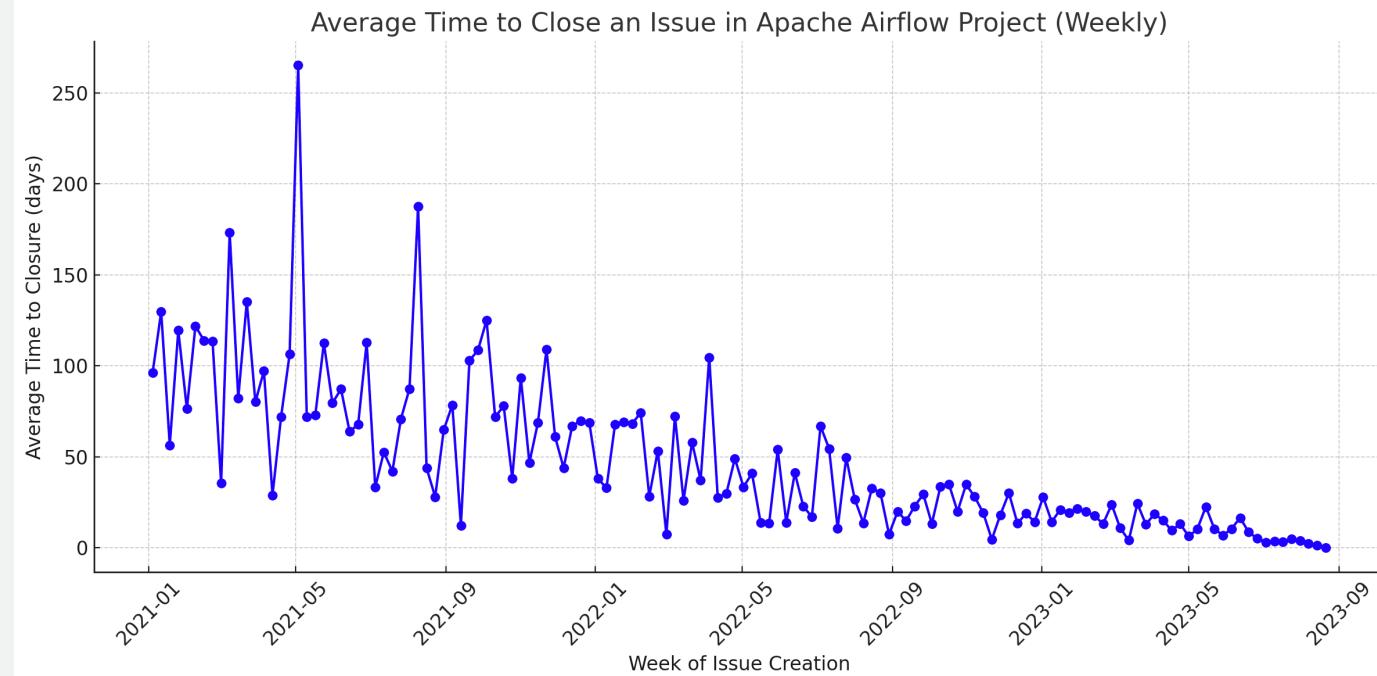
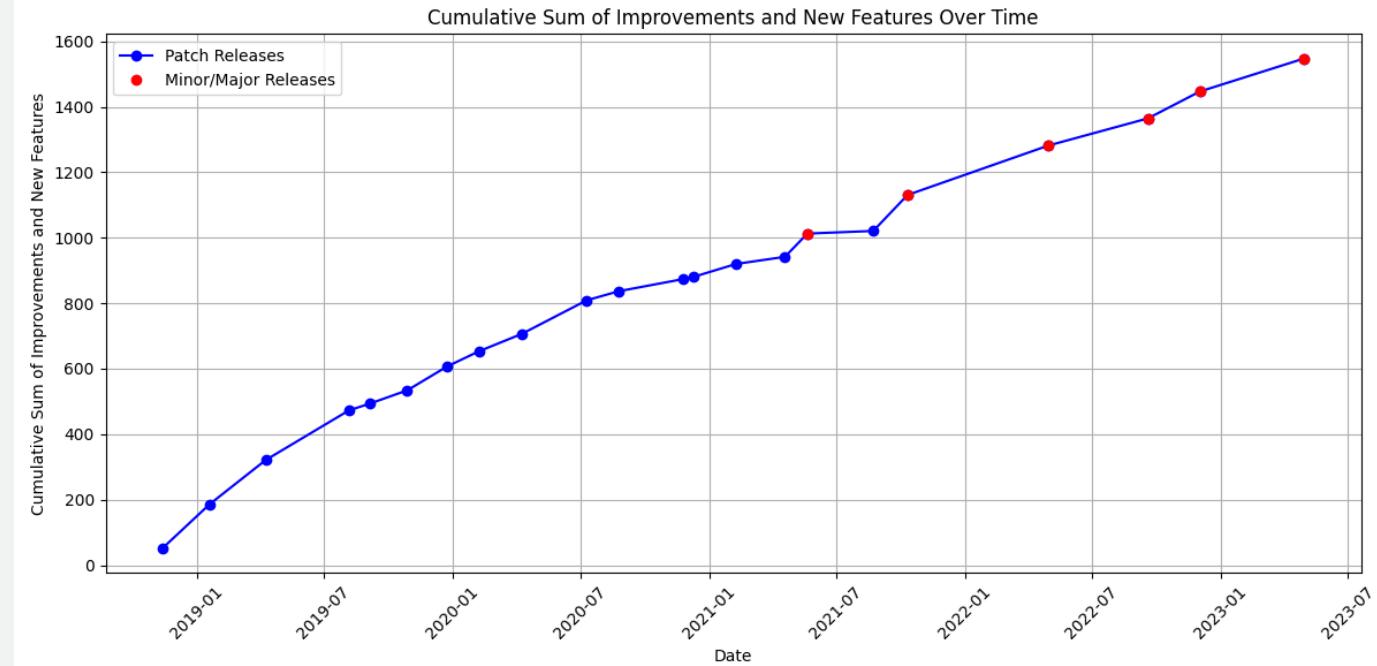


Adoption

- 10M monthly downloads
- 45% of data practitioners
- Apple, Shopify, Bloomberg

Roadmap

- 1600 new and improved features.
Avg 310 per year.
- New releases 30 days avg
- Time to close the bug < 10 days vs
100 days



Critical feedback 2 – Airflow is for batch oriented, time driven consistently running pipelines



Why critics claim?

- Time based scheduling
- Complex inter DAG/environment invocations
- Triggering DAG on external events

[Home](#) / [What is Airflow™?](#)

What is Airflow™?

Apache Airflow™ is an open-source platform for developing, scheduling, and monitoring batch-oriented workflows. Airflow's

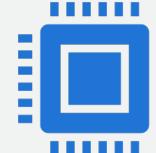
Dynamic workflow features



Data driven Scheduling



Deferrable operator/
Triggerer



Dynamic workflows



Micropipelines: A microservice approach for DAG authoring using datasets

by *Vikram Koka*

Reducing Cost with Async/Deferrable Operators

by *Zachary Bannor*



 **Airflow Summit**

Let's flow together

September 19-21, 2023,
Toronto, Canada

Task Duration

Task Tries

Landing Times

Gantt

Details



All Run Types

All Run States

Clear Filters

red failed queued removed restarting running scheduled shutdown skipped success up_for_rechedule

DAG
run_command

Details

Graph

More Details

DAG Runs Summary

Total Runs Displayed

1

Total success

1

First Run Start

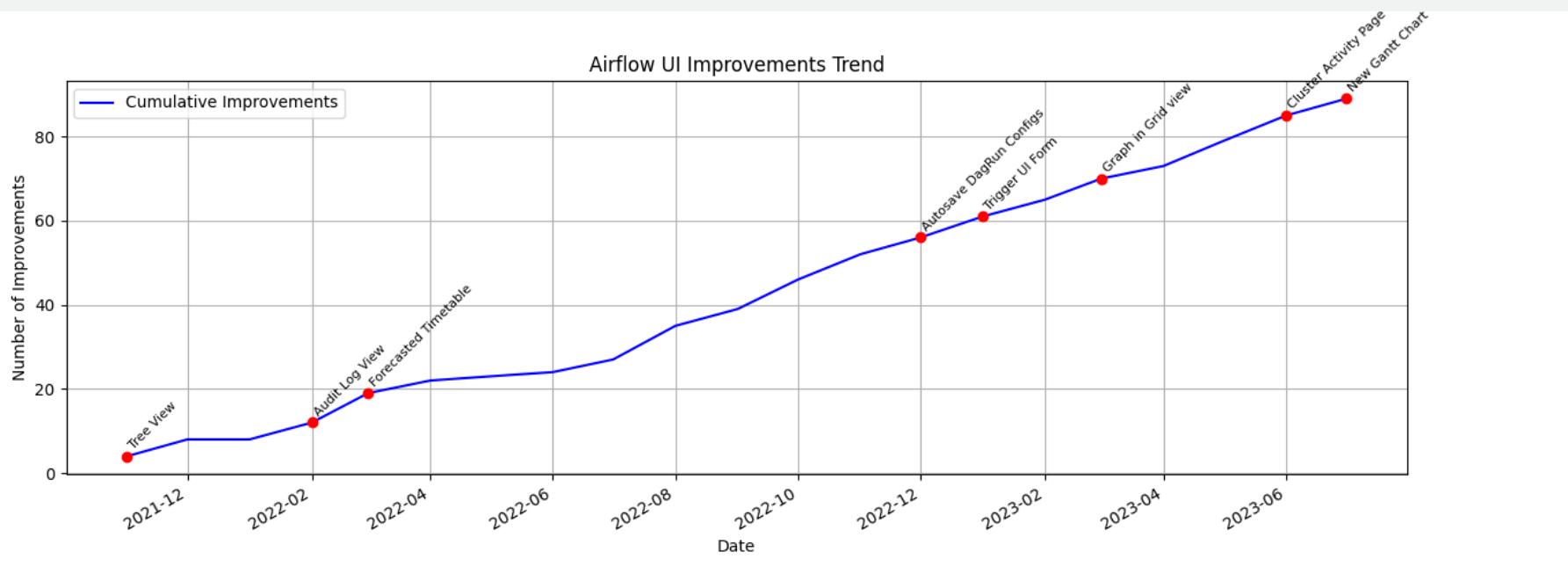
2023-05-10, 17:21:30

Last Run Start

2023-05-10, 17:21:30

Critical feedback 3 – Airflow UI is not intuitive

Airflow UI feature releases



Critical feedback 4 – Airflow offers limited decentralization



Criticism from Decentralization perspective

"People can potentially overwrite each other's DAGs. Credential management is complicated. Slow DAG can impact performance of whole Airflow. Getting DAGs to wait for each other (like one team prepares data up to a point and then other team builds on that) is kind of a nightmare. Etc etc."

-- Hacker News User

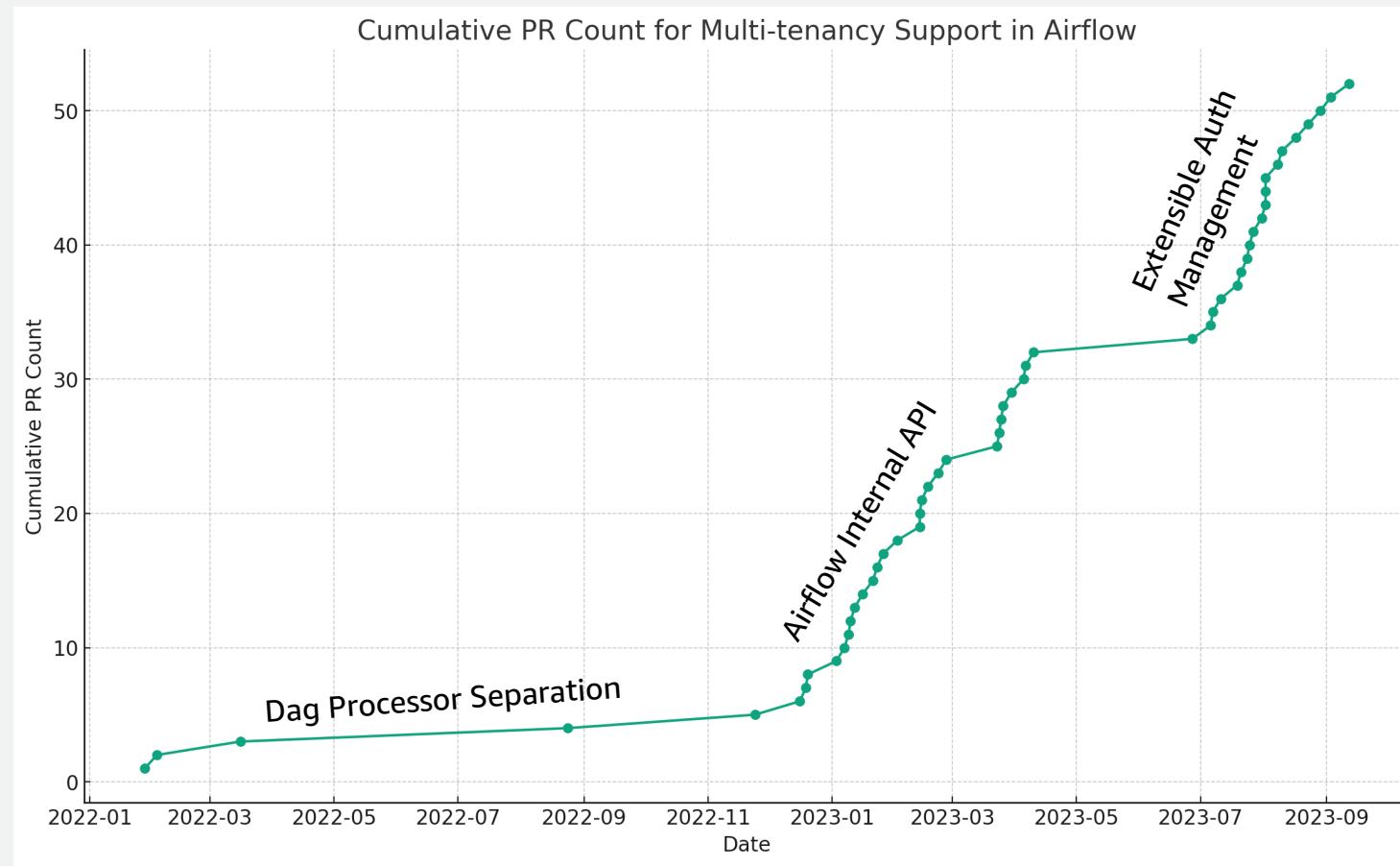


Security in multi-tenant deployment

Execution environment flexibility

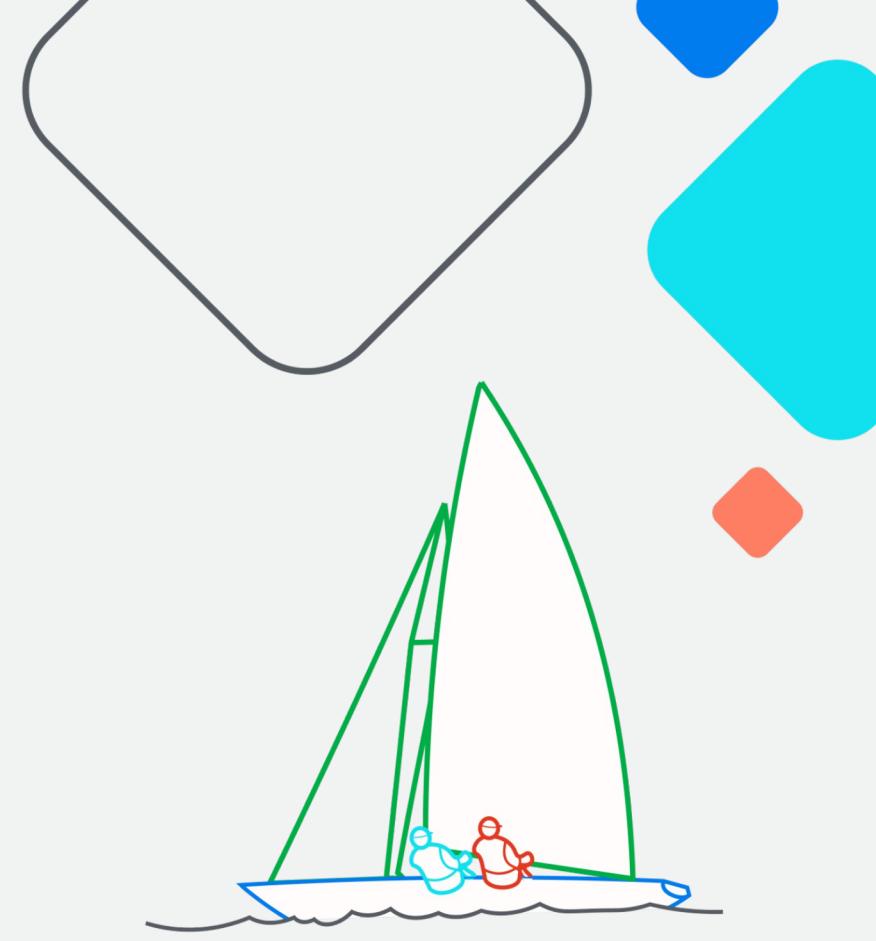
Auditing evolving workflows

Decentralization: Multi-tenancy Support



Multi-tenancy State of the Union

by Jarek Potiuk, Mateusz Henc & Vincent Beck



 **Airflow Summit**

Let's flow together

September 19-21, 2023,
Toronto, Canada

Decentralization: The path ahead for Airflow

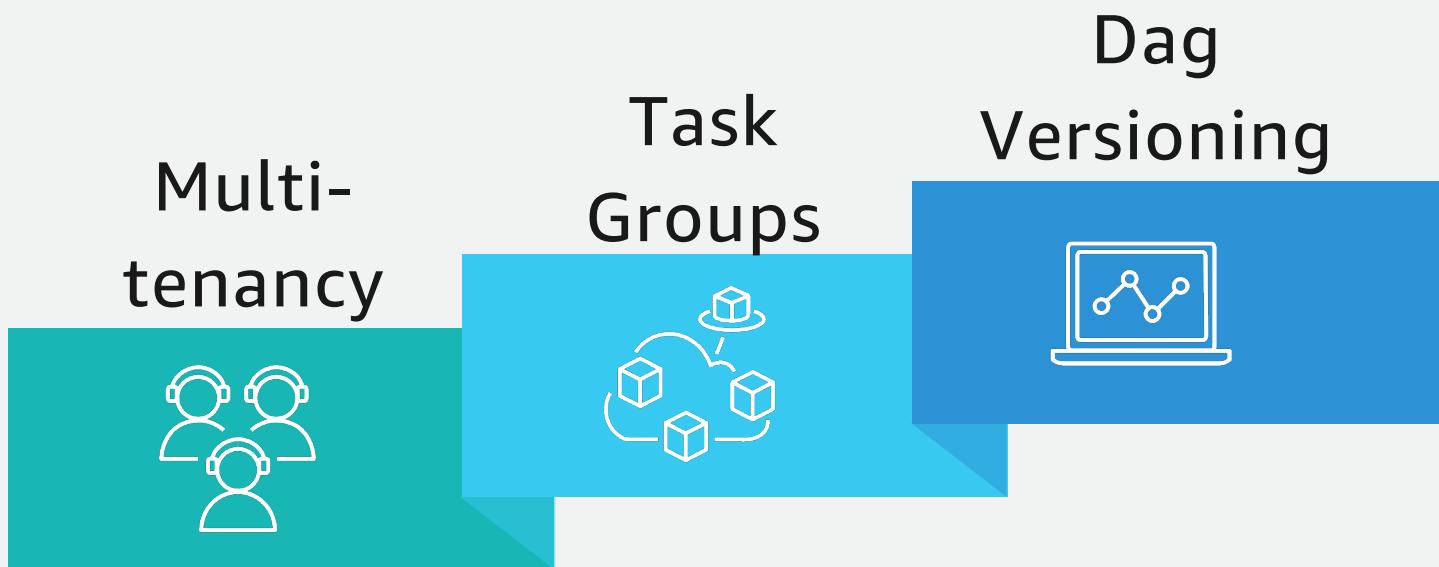
Multi-
tenancy



Task
Groups



Decentralization: The path ahead for Airflow

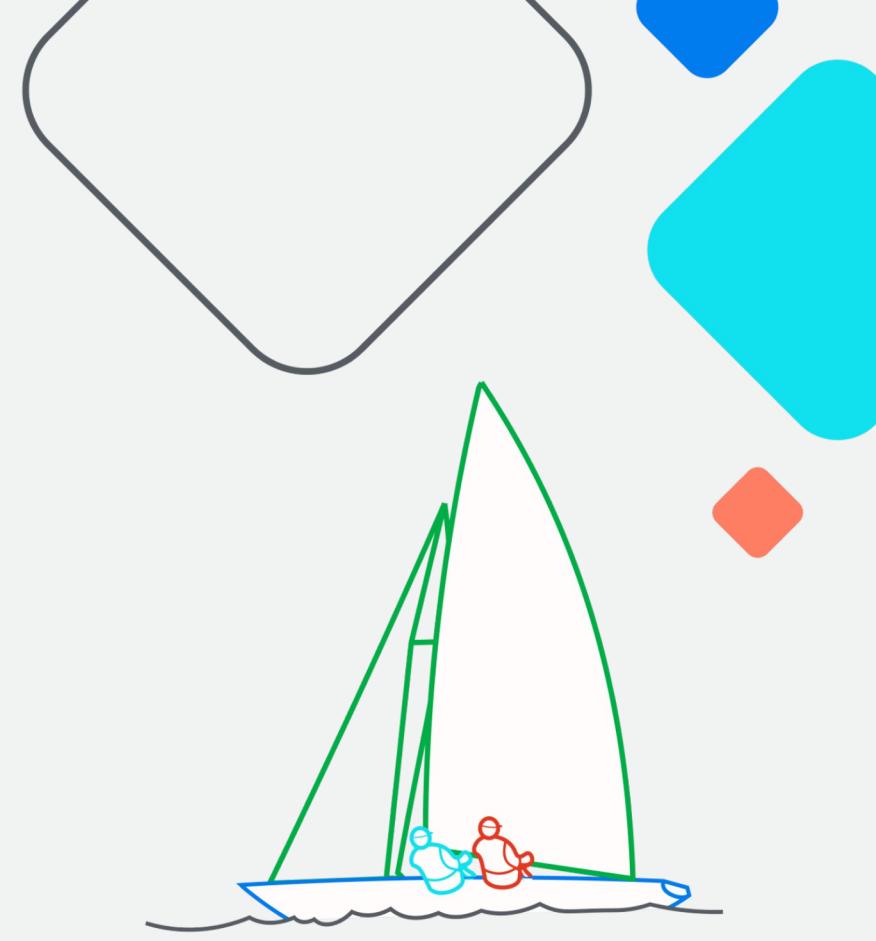


Decentralization: The path ahead for Airflow



Airflow Executors: Past, present and future

By Niko Oliveira

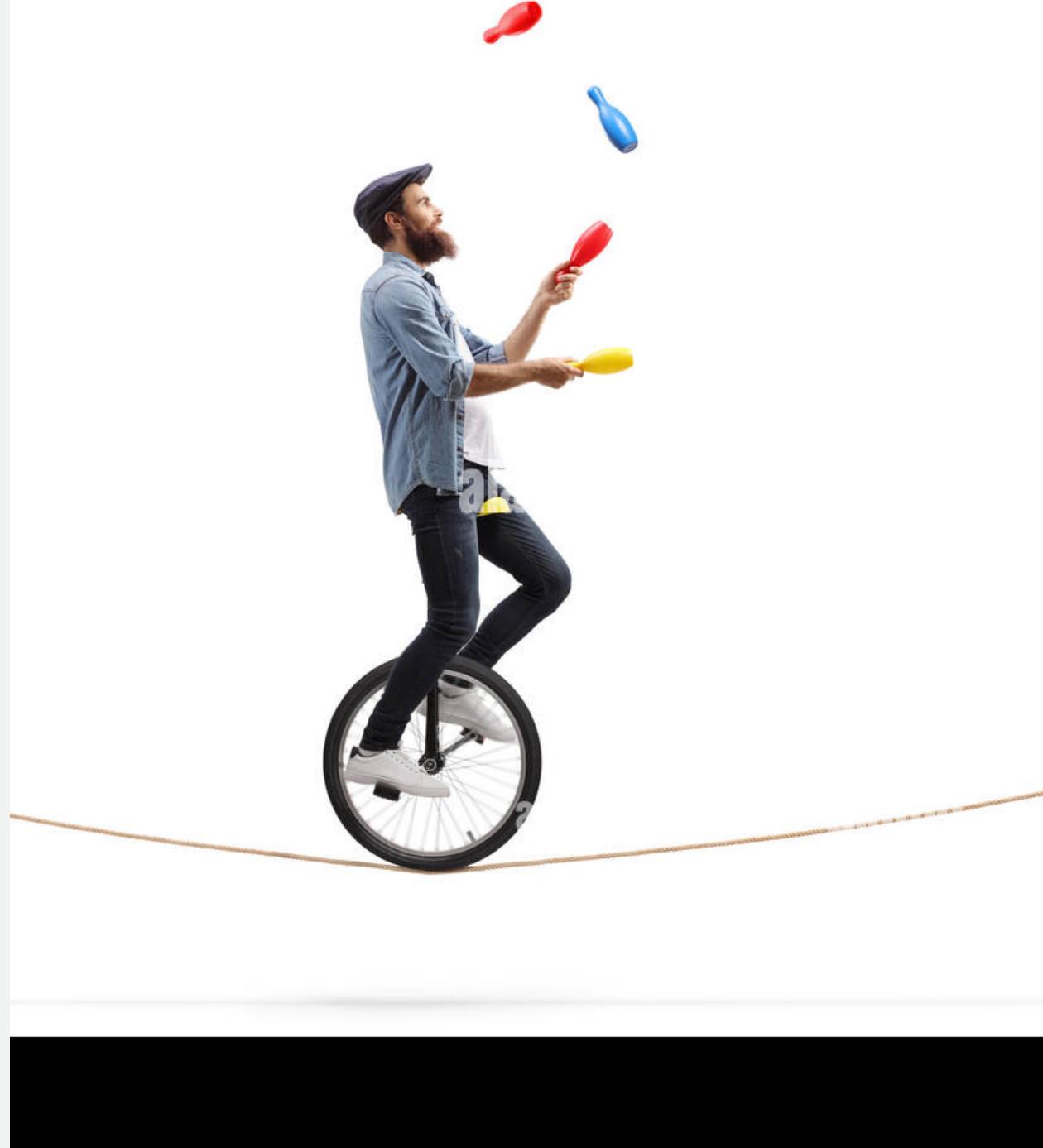


 **Airflow Summit**

Let's flow together

September 19-21, 2023,
Toronto, Canada

Critical feedback 5 – Airflow is hard to operate at scale



Why critics claim operating at scale is hard?

350+ Configurations

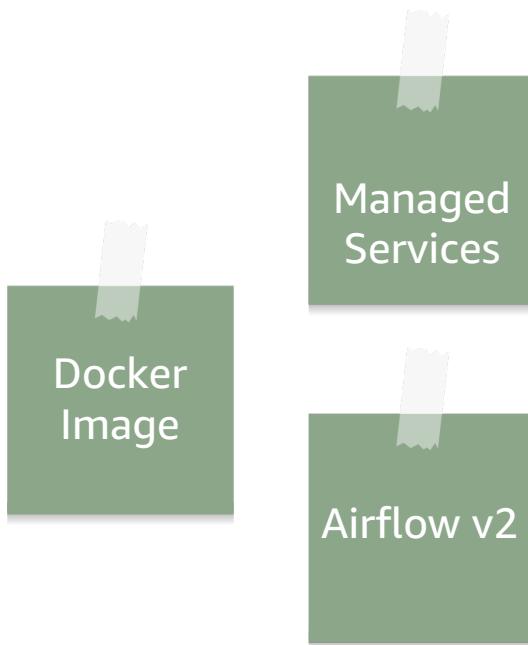
5+ Critical Components

Easy to break

"I used to be an Airflow fanatic as I loved the level of customisation you could have with it, but it's very cumbersome to manage at scale."
– Reddit user

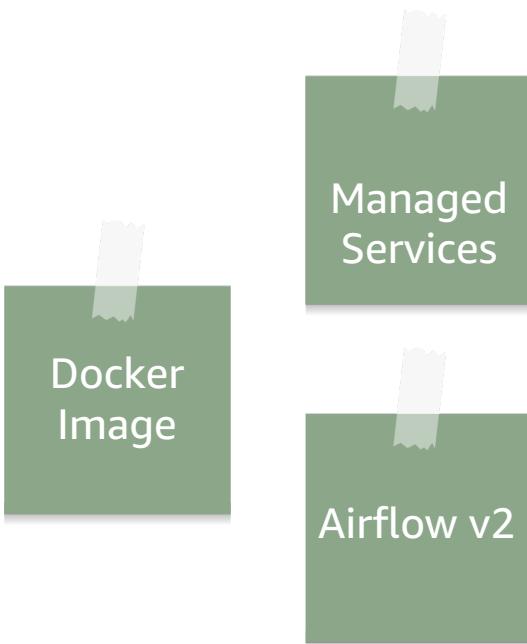
Navigating Airflow's Scaling horizons

Foundation



Navigating Airflow's Scaling horizons

Foundation

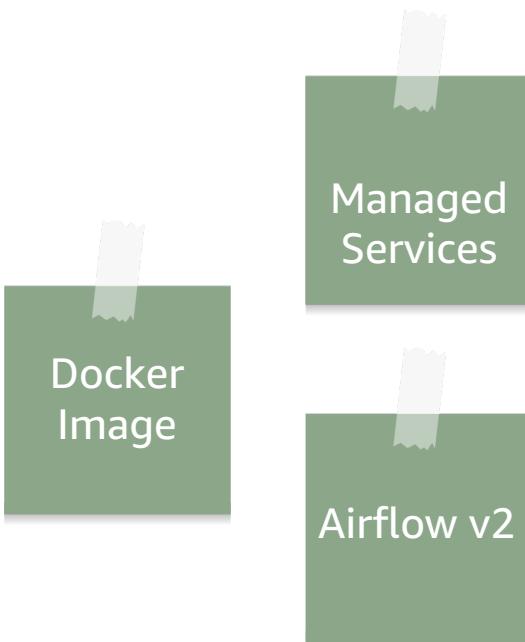


Adaptability



Navigating Airflow's Scaling horizons

Foundation



Adaptability



Innovation



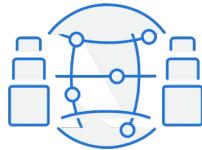
Critical feedback 6 – Airflow is not always developer friendly



Criticism about Developer Experience



Steep Learning Curve



Complex Dev-to-
Prod Integration



Hard to pinpoint
key metrics

*42% of users mentioned that the least discussed topic in Airflow contents is CI/CD implementation.**

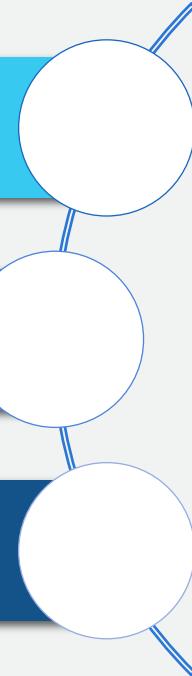


Developer-friendly road ahead for Airflow

Cluster Activity Dashboard

Simplified DAG Authoring

CI/CD Recipes & Plugins



44% of users indicated a preference for authoring DAGs using code alone. However, a significant portion, **40%**, expressed a preference for a hybrid approach combining both UI and code.*

What's Next?



2,606

GitHub Contributors



100,000+

Airflow Users



33,795

Slack Community

Getting Started



<https://airflow.apache.org/ecosystem/>

Ecosystem Page



<https://apache-airflow-slack.herokuapp.com/>

Slack Community

Thank You



Questions?

Shubham Mehta

LinkedIn: <https://www.linkedin.com/in/shubh-mehta>
X: https://x.com/smehta_

Uma Ramadoss

LinkedIn: <https://www.linkedin.com/in/uma-ramadoss>