dagster vs. Apache Airflow

Comparing the Open-source Solutions

Dagster and Apache Airflow are the two most popular generalized orchestrators used in data engineering today. Both are open-source solutions. For organizations looking to outsource the hosting and infrastructure concerns, the two leading solutions are Dagster+ by Dagster Labs and Astro by Astronomer.

Dagster vs. Airflow

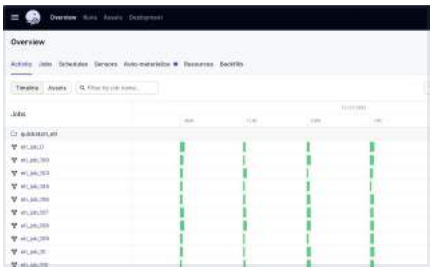
Before we consider the commercial hosting options, what truly separates these solutions are the core principles behind the designs.

Asset-oriented vs. Workflow-oriented Development

Dagster's asset-orientation makes it a central system of record for how your assets are defined and produced. Your team can collaborate on delivering critical data assets, not on the tasks of pipelines.

Airflow is workflow-oriented and task-centric. It does not provide asset-aware features or a coherent Python API. It is typically implemented after pipelines have been designed to trigger the required tasks.

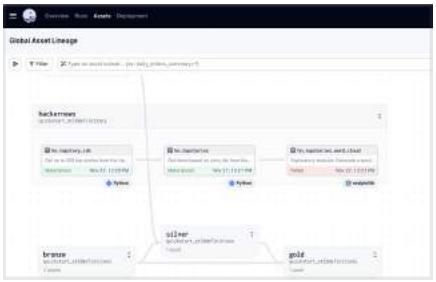
Testing and Debugging



Dagster is designed for use at every stage of the data development lifecycle. The open-source solution alone facilitates local development, unit testing, CI, code review, staging, and debugging. Dagster+ further expands these capabilities.

Airflow pipelines are harder to test and review outside of production deployments. Many teams working on Airflow end up doing their final testing in production. Here Astronomer helps provide options missing in the open-source solution.

Cloud-native Infrastructure



Dagster is cloud- and container-native, and designed for today's data infrastructure (ECS, K8s, Docker). Dependencies are easy to manage and upgrades are smooth. Dagster+ provides a turnkey hosting solution.

Isolating dependencies and provisioning infrastructure with Airflow is more complex and time-consuming than Dagster. Astronomer is a popular option for running Airflow as it helps mitigate these challenges.

dagster+ vs. ASTRONOMER

Comparing the Hosted Solutions

Airflow was introduced in 2014. Astronomer was founded in 2018 to augment Airflow to enhance developer productivity and optimize operational efficiency at scale. It has helped hundreds of organizations overcome the challenges of deploying and maintaining Airflow open-source.

Dagster Labs developed both Dagster open-source and Dagster+ as a new asset-oriented approach to data orchestration. Dagster Labs was founded in 2018. . By managing both the definition and the production of data assets, Dagster+ can leverage to rich metadata to provide novel features that go well beyond the execution of pipelines.

	Astronomer	Dagster+
Orientation of the framework	Workflow-oriented	Asset-oriented
Provider's contribution to the open-source project's code	55%	99%
CI/CD with branch deployments	Yes	Yes
Cost observability	No	Yes
Built in data asset cataloging	No	Yes
dbt model support	As tasks via Cosmos	Natively as data assets with rich metadata
Vendor lock-in alternative	Back to Airflow OSS or deploy on less feature-rich options like MWAA	Self-host Dagster OSS

dagster+ vs. ASTRONOMER

Organizations should select their orchestration solution based on their long-term vision for their data platforms.

Airflow is an established orchestration solution, and Astronomer addresses many of the foundational challenges related to operating Airflow, especially around deployment and maintenance. It is a good solution for organizations looking to remain on the Airflow workflow-oriented paradigm.

Dagster and Dagster+ are rapidly evolving to push beyond the traditional boundaries of orchestration. It adopts a contemporary Asset-oriented approach, building a system-of-record for critical data assets, adding new capabilities to the data platform such as cost observability, data cataloging, data quality tests and native dbt support.

Feature comparison

The above notwithstanding, here is a comparison table of features between Dagster+ and Astronomer

	Airflow	ASTRONOMER	dagster	dagster+
Core Focus	Workflow Orchestration	Workflow Orchestration	Data Orchestration	Data Orchestration
Primary Building Block	Tasks	Tasks	Assets	Assets
Safe cross-team collaboration	NO	NO	YES	YES
Partitioned data support	Limited	Limited	YES	YES
Sensors isolated from runtime	NO	NO	YES	YES
Cost observability	NO	NO	YES	YES
Basic Alerting	YES	YES	YES	YES
Conditional alerting	NO	Limited	YES	YES
Native data quality support	NO	NO	YES	YES
Environment management	NO	YES	YES	YES
Commercial Alternative hosting options	YES	NO	NO	NO