

The background features several stylized line art illustrations. At the top left is a telescope. At the top right is a wide-field instrument or camera. At the bottom left is a classical building with columns. At the bottom center is a dome-topped building. At the bottom right is a castle or fortress with a flag. The background is dark with various circles and stars representing celestial bodies.

AstroFlow Update

Brendan O'Brien

Royal Observatory of Edinburgh

brendan.obrien@roe.ac.uk

WHERE WE WERE (CORUÑA)

Moved to k8s

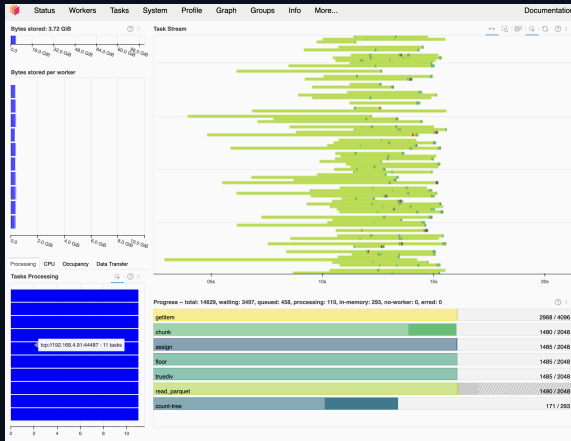


Moved to k8s
Replicated functionality



/ demo /	
browser	
Name	Modified
 1. Start here..ipynb	14 days ago
 2. Data holdings..ipynb	4 days ago
 3. Source counts over the sky....	4 days ago
 4. Mean proper motions over t...	4 days ago
 5. Working with Gaia XP spect...	14 days ago
 6. Working with cross-matche...	4 days ago
 7. Good astrometric solutions ...	4 days ago
 8. Tips and tricks..ipynb	14 days ago

Moved to k8s
Replicated functionality
Added dask



```
> 2024-11-18 16:00:27.035 {"time": "2024-11-18T16:00:27+00:00", "remote_addr": "", "x_forwarded_for": "", "request_id": "8e0449f0050767a964582cf58426f4a1", "remote_user": "", "bytes_sent": 3226, "request_time": 0.012, "status": 200, "vhost": "hub.arcus-demo.astro-flow.com", "request_proto": "HTTP/2.0", "path": "/user/24969d2a-dbd-4ffd-a06f-386143fcd53/api/contents/demo", "request_length": 120, "duration": 0.012, "method": "GET", "user_agent": "Mozilla/5.0 (Macintosh; Intel Mac OS X 10.15; rv:128.0) Gecko/20100101 Firefox/128.0", "upstream_name": "192.168.173.214:8000", "upstream_addr": "192.168.173.214:8000", "upstream_status": "200", "req_id": "8e0449f0050767a964582cf58426f4a1", "req_body": "", "req_content_type": "application/json", "res_content_type": "application/json", "upstream_response_length": "2919", "upstream_response_time": "0.012", "cache_status": "", "upstream_bytes_sent": "1948", "upstream_bytes_received": "3340"}
```

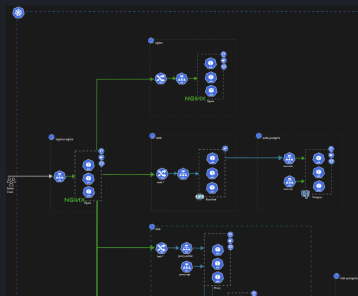
Added Observability & Standardised Logging



Moved to k8s
Replicated functionality
Added dask
Added Observability & \$
Added Documentation

AstroFlow Architecture

The overall cluster architecture is shown below.



Architecture

This page details the architecture of AstroFlow. It refers to components listed [here](#)

K8s Architecture

The cluster architecture is detailed in the following sections, starting with the overall cluster architecture before drilling down into areas separated by logical functionality/namespaces. These are described in terms of kubernetes native objects.

Description of k8s components

The entry point to the cluster for public traffic is via a cloud provider load balancer with a publically assigned IP address, on a suitable subnet, with appropriate ports exposed via security groups etc. All of this if configured in the cloud provider tooling during cluster initialisation and is described [here](#).

Note

Public traffic means a traffic ingress with a public IP address, it does not imply there is general public access to the product. The platform implements strong authentication and authorization.



Moved to k8s

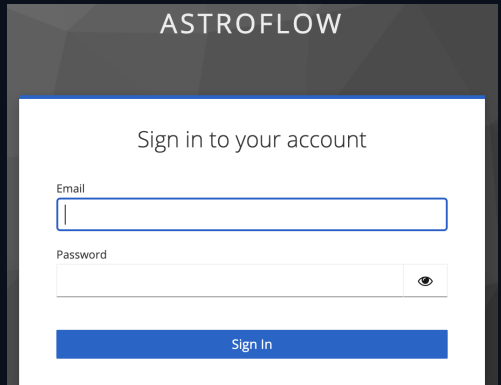
Replicated functionality

Added dask

Added Observability & Standardised Logging

Added Documentation

Single Sign On



The image shows a screenshot of the ASTROFLOW login interface. At the top, the word "ASTROFLOW" is displayed in white capital letters on a dark grey background. Below this, the text "Sign in to your account" is centered. There are two input fields: "Email" and "Password". The "Email" field is a simple text box with a vertical cursor. The "Password" field is a text box with a toggle icon (an eye) on the right side. Below the password field is a blue button labeled "Sign In".

ASTROFLOW

Sign in to your account

Email

Password

Sign In

Moved to k8s

Replicated functionality

Added dask

Added Observability & Standardised Logging

Added Documentation

Single Sign On

Threat detection

08:33

74%

←

Alert #267

...

Detail

Notes

Logs

Responder states

P1

OPEN

06:32:07.548562629: Warning

Sensitive file opened for reading

by non-trusted program (file=/

etc/shadow gparent= ggparent=

#267

x1

Sep 27, 2024 7:32 AM (GMT+01:00)

No Owner

Also ...

Also ... Gandalf

The screenshot shows a Jupyter Notebook interface in a web browser. The browser address bar displays the URL: `hub.arcus-demo.astro-flow.com/user/247d4c07-a759-4e0f-b913-8838237e7c02/lab/work...`. The notebook is titled "1-gandalf-train.ipynb" and is running on "Python 3 (ipykernel)". The left sidebar shows a file explorer with a directory structure: `/ demo / gandalf /`. The files listed are: `data`, `images`, `0-apogee-eda.ip...`, `1-gandalf-train.ip...` (selected), `2-gandalf-test.ip...`, `3-generative-fun...`, `4a-bokeh-ganda...`, `4b-bokeh-ganda...`, `README.md`, `train_config.yaml`, and `visualisation_con...`. The main content area of the notebook displays the "SPACIOUS" logo, which is a stylized, pixelated representation of the word. Below the logo, there is a banner for the "SPACIOUS" project, funded by the European Union, with the text: "SPACIOUS is a Horizon Europe HORIZON-CL4-2023-SPACE-01-71 project funded under grant agreement no. 101555205". The URL `https://spacious.ub.edu/` is also visible. Below the banner, the text "GANDALF 3.2.0" is displayed. The main title of the notebook is "GANDALF Train tutorial". At the bottom, the first line of code is shown: `[ ]: from gandalf import GandalfTrain`. The text "Arguments and GANDALF class" is visible at the bottom of the notebook.

hub.arcus-demo.astro-flow.com/user/247d4c07-a759-4e0f-b913-8838237e7c02/lab/work...

1-gandalf-train.ipynb

notebook is read-only Python 3 (ipykernel)

SPACIOUS

Funded by the European Union
SPACIOUS is a Horizon Europe HORIZON-CL4-2023-SPACE-01-71 project funded under grant agreement no. 101555205
<https://spacious.ub.edu/>

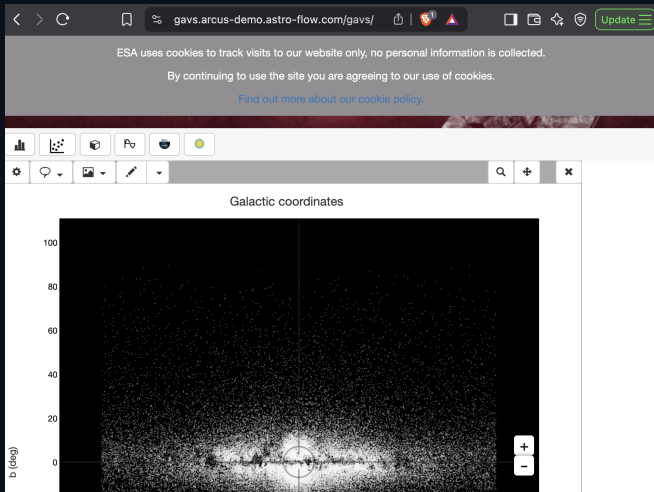
GANDALF 3.2.0

GANDALF Train tutorial

```
[ ]: from gandalf import GandalfTrain
```

Arguments and GANDALF class

Also ...
Gandalf
GAVS



Also ...

Gandalf

GAVS

Testing

Also ...

Gandalf

GAVS

Testing

Production Readiness Plan

Also ...

Gandalf

GAVS

Testing

Production Readiness Plan

Multiple environments



Google Cloud



openstack™

Also ...

Gandalf

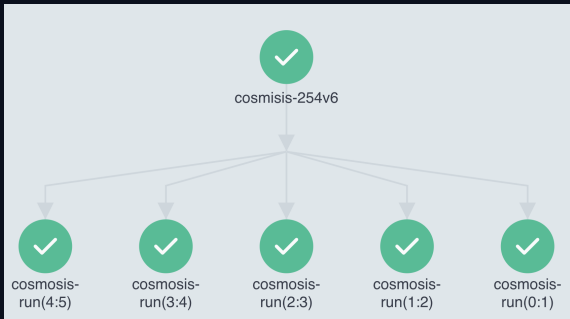
GAVS

Testing

Production Readiness Plan

Multiple environments

CLOE/Cosmosis



WERE WE ARE NOW

Testing Pipelines

← wfau / astroflow

Tasks

Config

Changed files 2

Debug

 brendanobrien-roe [feature/#139](#) [3ef95a0e0a](#)

✓ clone	01:14
✓ init	00:03
✓ inflate	00:11
✓ kube-score	00:11
✓ secret-leaks	00:14
✓ unit-test-portal	00:23
✓ upload-reports	00:04

Step Logs

```
1 + git
2 Init:
3 + git
4 + git
5 + git
6 From
7 * bl
8 + git
9 + git
10 + git
11 Fetch
12 + git
```

Testing Pipelines

← wfau / astroflow

Tasks

Config

Changed files 2

Debug

 brendanobrien-roe [feature/#139](#) [3ef95a0e0a](#)

Step Logs

✓ clone	01:14
✓ init	00:03
✓ inflate	00:11
✓ kube-score	00:11
✓ secret-leaks	
✓ unit-test-portal	
✓ upload-reports	00:04

```
1 + git
2 Init:
3 + git
4 + git
5 + git
6 From
7 * b1
8 + git
9 + git
10 + git
```

```
k8s/tests/test_build_volumes.py::test_basic_pv_and_pvc PASSED
k8s/tests/test_build_volumes.py::test_get_persistent_volume_cla
k8s/tests/test_build_volumes.py::test_get_persistent_volume_cla
```

Testing Pipelines

Guasom & Full Coruña integration (inc. docs)

[Home](#) [User Guide](#) [Data Holdings](#) [Analysis Applications](#)

Analysis Applications

Gandalf

API

Train

Test

Visualiser

SOM

GUASOM

GANDALF

Generative Adversarial Networks for Disentangling And Learning Framework (GANDALF)



GANDALF

GANDALF is an ad-hoc framework written in Python to define, train, test, and visualize disentangling models. Disentangled learning is a form of unsupervised learning focused on identifying and isolating the underlying independent factors of variation within the data.

The framework is divided into three main parts. The first, for train, is the `GandalfTrain` class. The second is the `test` module, with functions to test the trained models. Finally, the `visualizer` module allows us to explore the data with an interactive bokeh visualizer.

Table of contents

Train

Set up

Train process

Additional Resources

Test

t-SNE

Conditional parameters analysis

Additional Resources

Visualizer

Configuration

Bokeh server

Jupyter Notebook

Additional Resources

Testing Pipelines

Guasom & Full Coruña integration (inc. docs)

Full Lisbon Integration (inc. cluster access)

Testing Pipelines

Guasom & Full Coruña integration (inc. docs)

Full Lisbon Integration (inc. cluster access)

New environments → RAL (Euclid)

Testing Pipelines

Guasom & Full Coruña integration (inc. docs)

Full Lisbon Integration (inc. cluster access)

New environments → RAL (Euclid)

Production quality

Testing Pipelines

Guasom & Full Coruña integration (inc. docs)

Full Lisbon Integration (inc. cluster access)

New environments → RAL (Euclid)

Production quality

Migrated all users

Testing Pipelines

Guasom & Full Coruña integration (inc. docs)

Full Lisbon Integration (inc. cluster access)

New environments → RAL (Euclid)

Production quality

Migrated all users

Access to early release DR4 PVP data via iceberg mechanism

Testing Pipelines

Guasom & Full Coruña integration (inc. docs)

Full Lisbon Integration (inc. cluster access)

New environments → RAL (Euclid)

Production quality

Migrated all users

Access to early release DR4 PVP data via iceberg mechanism

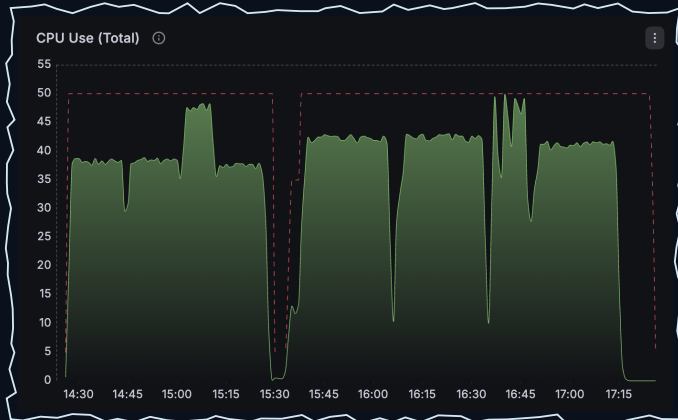
Some new features e.g. user docs and kyverno policy management

ONGOING DEVELOPMENT

Workload isolation, quotas etc

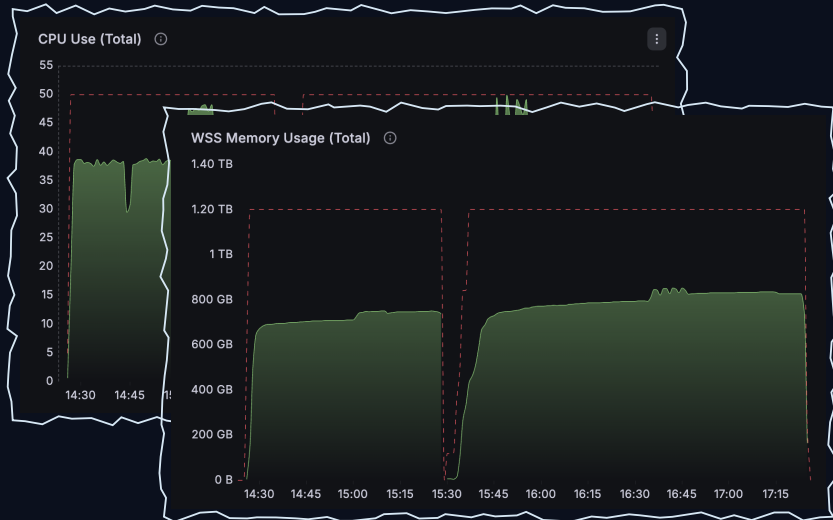
Workload isolation, quotas etc

User-level resource reporting



Workload isolation, quotas etc

User-level resource reporting



Workload isolation, quotas etc

User-level resource reporting

Making provisioning more dynamic

Kubernetes

User's Service Account:

ENVIRONMENT	NAME	DIRECT	GROUP
somerville-uat	user-{username}	False	/jupyterhub

User's Quota:

ENVIRONMENT	NAME	CPU LIMIT CORES	MEMORY LIMIT GiB	DIRECT	GROUP
somerville-uat	Basic Jupyterhub Quota	100	100	False	/jupyterhub

User's PVCs:

PVC	PV	HOST PATH	MOUNT PATH	DIRECT	GROUP
pvc-gaia-dr3-data-ceph	pv-gaia-dr3-data-ceph-user-{username}	/mnt/GAIA_DR3_ALL_BUCKETED_2048_RO	/mnt/gaia-dr3-data-ceph	False	/jupyterhub
pvc-gaia-edr3-data-ceph	pv-gaia-edr3-data-ceph-user-{username}	/mnt/GAIA_EDR3_BUCKETED_2048_RO	/mnt/gaia-edr3-data-ceph	False	/jupyterhub
pvc-gaia-dr3-data-ssd	pv-gaia-dr3-data-ssd-user-{username}	/mnt/dr3_data_share	/mnt/gaia-dr3-data-ssd	False	/jupyterhub
pvc-gaia-edr3-data-ssd	pv-gaia-edr3-data-ssd-user-{username}	/mnt/edr3_data_share	/mnt/gaia-edr3-data-ssd	False	/jupyterhub
pvc-iceberg-playground-ceph	pv-iceberg-playground-ceph-user-{username}	/mnt/ICEBERG_PLAYGROUND_SHARE	/mnt/iceberg-playground	False	/iceberg
pvc-scratch	pv-scratch-user-{username}	/mnt/ssd_cache	/mnt/scratch	False	/jupyterhub
pvc-vvxx-processed-data-ceph	pv-vvxx-processed-data-ceph-user-{username}	/mnt/VVXX_PROCESSED_DATA_SHARE	/mnt/vvxx-processed-data-ceph	False	/vvxx
pvc-vvxx-source-data-ceph	pv-vvxx-source-data-ceph-user-{username}	/mnt/VVXX_DATA_SHARE	/mnt/vvxx-source-data-ceph	False	/vvxx

Workload isolation, quotas etc

User-level resource reporting

Making provisioning more dynamic

Iceberg metadata work

Workload isolation, quotas etc

User-level resource reporting

Making provisioning more dynamic

Iceberg metadata work

Feedback & signposting

Experience running in Production ...

Adding users

Adding users

Self registration

* Required fields

Register

Email *

Password *

☐

Confirm password *

☐

First name *

Adding users

Self registration

Approval

ID *

247d4c07-a759-4e0f-b913-8838237e7c02

Created at *

10/30/2024, 6:19:24 PM

Required user actions



Select action

Email verified 



On

Adding users

Self registration

Approval

Access management

☐	Group membership	Path
☐	argo-ci	/argo-ci
☐	bsc-gaia-pvp	/bsc-gaia-pvp
☐	docs	/docs
☐	euclid	/euclid
☐	iceberg	/iceberg
☐	jupyterhub	/jupyterhub
☐	monitoring	/monitoring
☐	performance-testing	/performance-testing
☐	udc	/udc
☐	user-docs	/user-docs

Adding users

- Self registration

- Approval

- Access management

Requests to add packages

- Less frequent but easy to deal with

INFRASTRUCTURE ISSUES BEYOND OUR CONTROL



Emergency patches

Power issues

Scheduled maintenance

→ mostly automatic recovery

Emergency patches

Power issues

Scheduled maintenance

→ mostly automatic recovery

Storage attach/detach issues

Corruption issues

→ mitigation strategy

Virtualised storage ...

OpenStack Block Storage

A few large OpenStack volumes provide storage to the cluster.



...

Longhorn Master Volumes

Each OpenStack volume is used as a Longhorn master volume (disk).



...

Longhorn Logical Volumes

Each master volume is split into multiple logical volumes for Kubernetes workloads (PVCs).



...



...



...

Replicas across Kubernetes nodes

Each logical volume has multiple replicas (e.g. 3), with all replicas stored on different Kubernetes nodes using anti-affinity.

Kubernetes Node 1



Kubernetes Node 2



Kubernetes Node 3



...

Less attach/detach events

→ less metadata corruption

Abstracts away the underlying storage type

Backup/restore is easy

Redundancy

→ One disk drops → recover from another

WHAT'S NEXT

Finish the various on-going developments

Continue integrations

Continue to seek and implement feedback

Scaling

SPACIOUS-NET