



Updates in the AI4OS stack

AI4EOSC collaboration
iheredia@ifca.unican.es
heredia@ifca.unican.es
obregonm@ifca.unican.es
sftobias@ifca.unican.es

20th March 2024
iImagine User Workshop



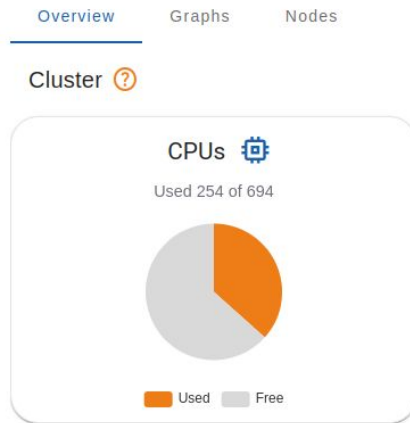
Outline



Dashboard







New features

- Enables users to interact with the platform in a simple and graphic way.
- It is divided into: (1) Marketplace and (2) Deployments
 - (1) Allows users to launch deployments and tools.
 - (2) Displays the current deployments of the user.
- *Upcoming features:*
 - **Secrets:** Tools Integration with Vault. Users can create and delete secrets that will be available inside their deployments.
 - **Stats:** authorized clients will be able to check the stats of the platform (cluster stats, user's aggregate per day, use over time and nodes stats).
 - **OSCAR** integration: easily connect deployments with running oscar instances.




TestFL

List of secrets

Developers			
default			

Items per page: 5 1 - 2 of 2 << < > >>

New secret name*

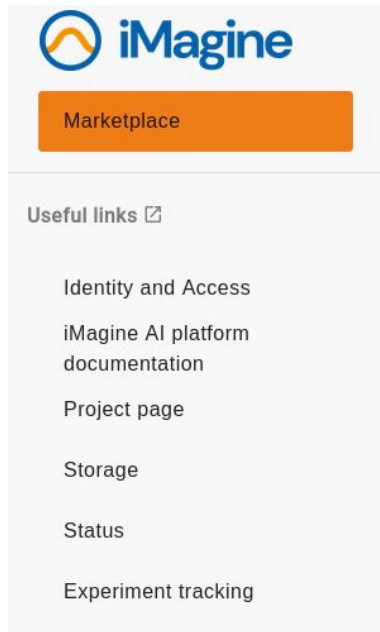
Add 

Close

Federated Learning server

New features

- Currently available as a tool in the dashboard
- Goal: train ML/DL models in a distributed way without sharing data from different data owners (clients), thus ensuring privacy
- Allows to perform a FL training with the data from different clients without leaving their own local or cloud machines
- Deployed using [Flower](#)
- Tutorials available in the AI4OS docs
- *Upcoming features:*
 - Clients will use a token for connecting with the server (ensuring that only authenticated clients can participate in the training)
 - Secret management using [Vault](#)



Marketplace 

Modules

Tools

 Federated learning server

Federated learning server with Flower

- Currently offers *concept and data drift* detectors that enable the detection of significant changes in concepts previously learned by the model or alterations at the feature distribution level, which may impact the model's performance. This includes:
 - 13 *concept drift* detectors.
 - 18 *data drift* detectors.
- Incorporates the concept of *callbacks*, a popular feature in libraries such as PyTorch Lightning and Keras. This enables users to execute custom code at key stages during the detector's operation.
- Include datasets and metrics for evaluating the detectors' performance.
- *Upcoming features*:
 - Regular additions of new detectors.
 - Introducing new *callbacks* to address a wider range of real-world use cases.
 - Enhancing the performance of *concept drift* detectors through the utilization of Cython, and adapting *data drift* methods for GPU utilization.

<https://docs.ai4os.eu>

- New TOC
- New quickstart
- New Frequently Asked Questions (FAQ) section
- Tutorials for new tools:
 - MLflow
 - Elyra
 - Flowfuse
- New video tutorials playlist

Frequently Asked Questions (FAQ)

This page gathers known issues of the platform, along with possible solutions. If your issue does not appear here, please contact support.

🔥 Service X is not working

Check the [Status page](#) to see if there is any maintenance action going on. If you don't see anything, wait a couple of hours to make sure it is not a temporary issue.

If the issue persists, please contact support.

🔥 The Dashboard says I only have 500 MB of disk in my deployment

In your deployment information, you might see that under `Disk memory` your deployment has 500 MB assigned which is much less than what you might have asked initially.

For the time being, this number is meaningless, because we are not enforcing correctly the disk limits. Users have access to all the resources of the node, and they might conflict with other users disk space. This is why we kindly ask users to respect a **maximum** of 20 GB of disk usage per deployment.

We are planning to fix this issue in the new cluster we are setting up.

If you need more than 20 GB, please check the [provided option](#) of accessing your dataset via a virtual filesystem, in order to avoid overloading the disk.

🔥 I ran out of disk in my deployment

The current Nomad cluster has not the ability to properly isolate disk between different users using the same physical machine. So it might be the case that some user might be using more resources than their due share thus consuming the disk of other users that share their same node. We are planning to fix this issue in the new cluster we are setting up.

First, make sure to delete files in the Trash (`/root/.local/share/Trash/files`). Files end up there when deleted from the JupyterLab UI, thus not freeing up the space correctly.

In the meantime, if you are sure that you are *using less than 20 GB of disk*, but you still find that there is not disk left, please contact support.

Nomad inference endpoints

Coming soon

- New short-lived endpoint to quickly test a model
- Deployable from the Dashboard
- Complementary to OSCAR endpoints
- 4 instances behind a load balancer

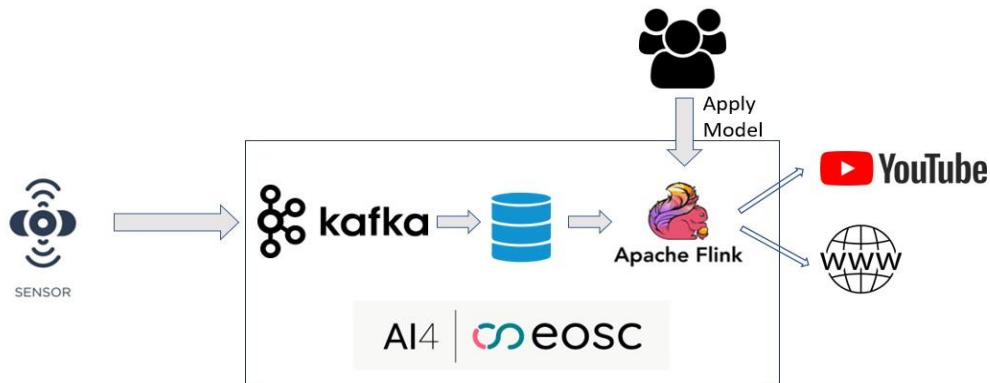
TRY IT



Kafka for video streaming

Coming soon

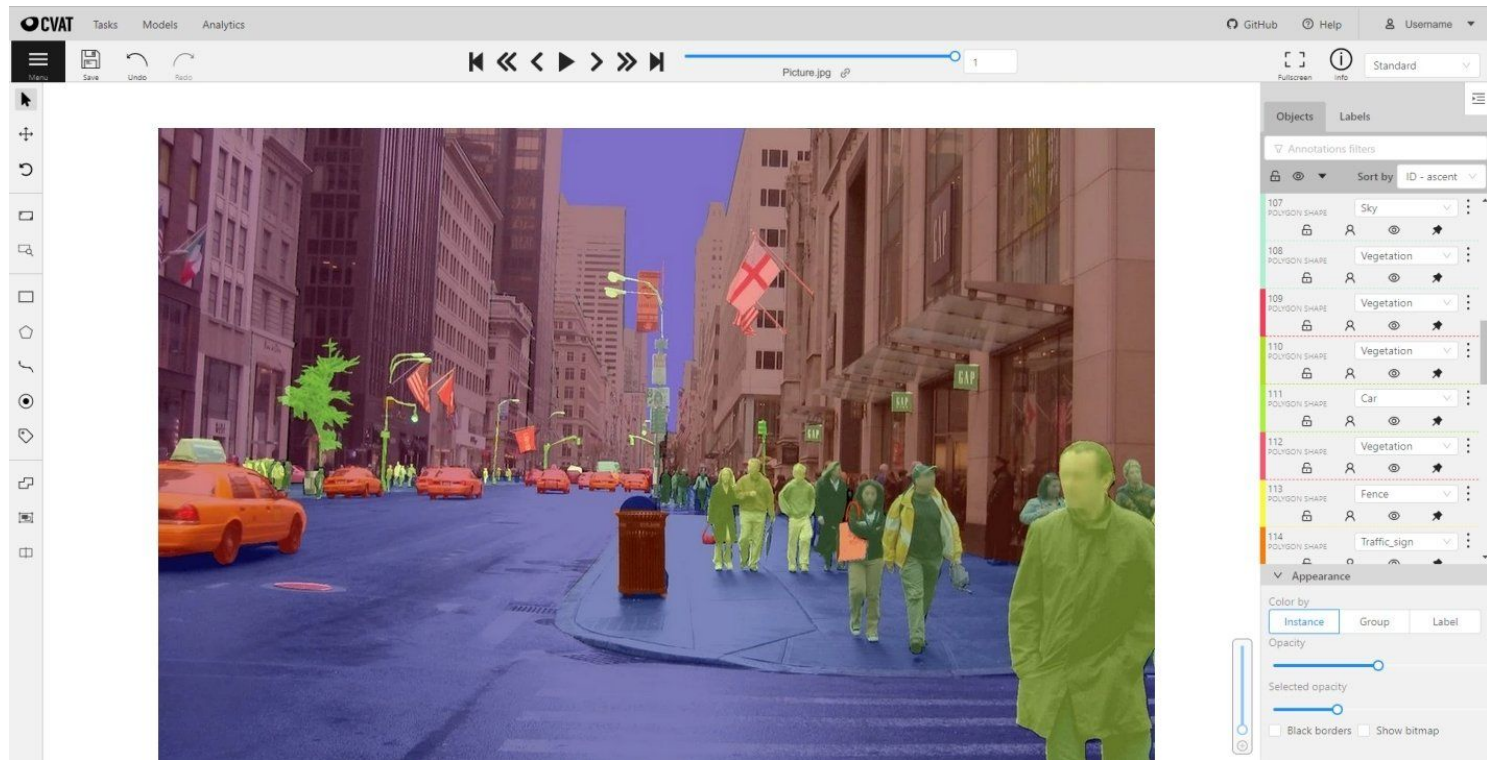
- Currently testing in the AI4EOSC platform
- Kafka collects the streaming data from different sensors, spreading the computing in multiple nodes
- Once Kafka is deployed inside the ai4eosc platform, an endpoint is available for applying processing (cleaning, ml models, etc.) to these streams in real time
- Processing tools will be deployed to automate the data processing
- *Upcoming:*
 - End testing and include the service in the dashboard
 - Connect Kafka with other processing tools like Apache Flink to generate the final stream for video streaming



CVAT tools for labelling

Coming soon

New tool for image labelling → ideally connected to Nextcloud



Upcoming features:

- Translation of metadata to RDF
- Additional metadata (Github commit hash, Docker image hashes, MLflow info)
- Custom graphs on the Dashboard with [Metaclip](#)
- Automatic assessment of FAIRness of data using [FairEVA](#) with custom plugin for iMagine data repository (Zenodo?).

