



Present and future state of the platform *(user oriented)*

Andrés Heredia (heredia@ifca.unican.es)
Ignacio Heredia (iheredia@ifca.unican.es)



AI4EOSC Platform: User's workshop (November 2023)

AI4

 eosC

Present

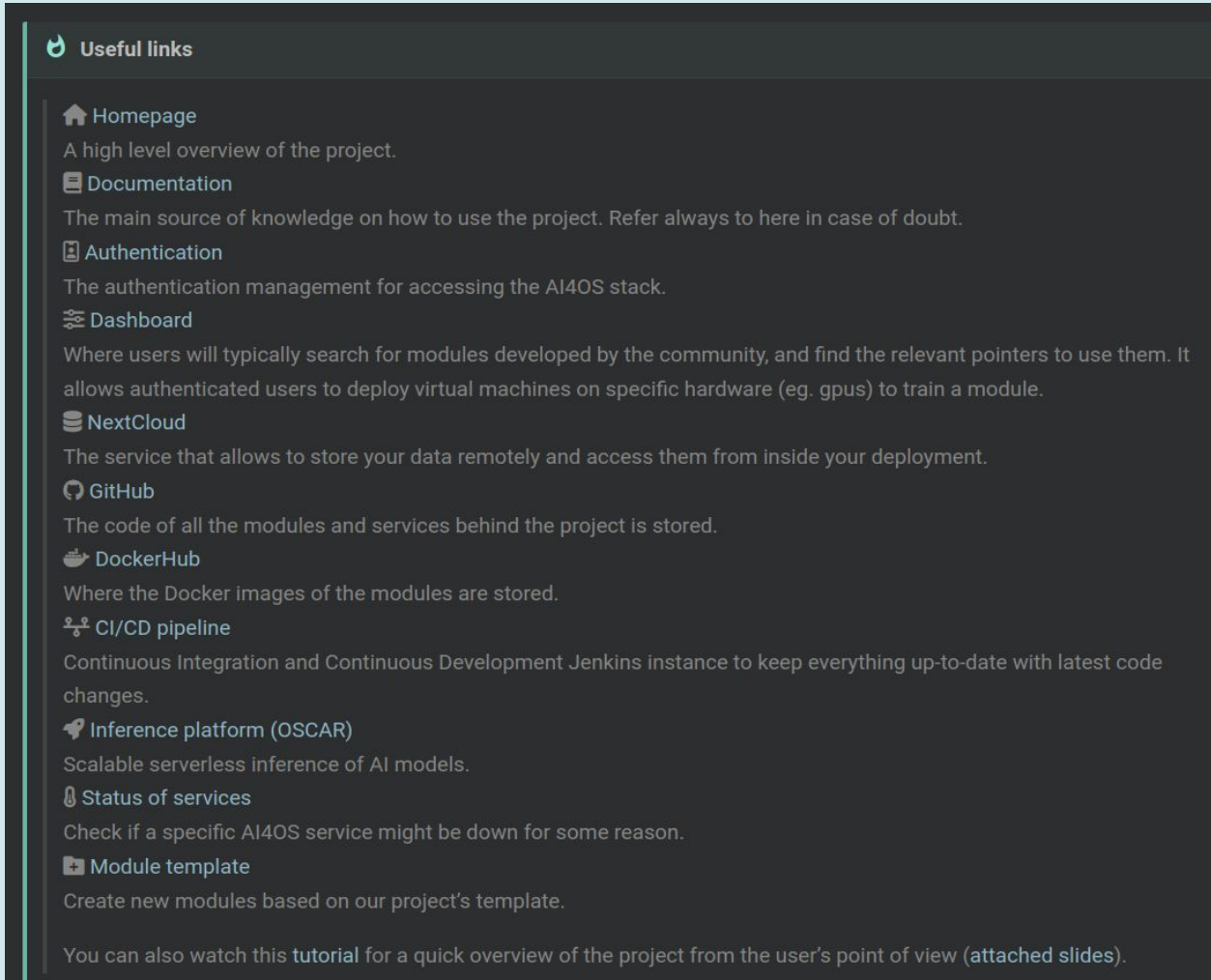
Future



Co-funded by
the European Union

General overview - Docs

<https://docs.ai4eosc.eu/>



Useful links

- Homepage**
A high level overview of the project.
- Documentation**
The main source of knowledge on how to use the project. Refer always to here in case of doubt.
- Authentication**
The authentication management for accessing the AI4OS stack.
- Dashboard**
Where users will typically search for modules developed by the community, and find the relevant pointers to use them. It allows authenticated users to deploy virtual machines on specific hardware (eg. gpus) to train a module.
- NextCloud**
The service that allows to store your data remotely and access them from inside your deployment.
- GitHub**
The code of all the modules and services behind the project is stored.
- DockerHub**
Where the Docker images of the modules are stored.
- CI/CD pipeline**
Continuous Integration and Continuous Development Jenkins instance to keep everything up-to-date with latest code changes.
- Inference platform (OSCAR)**
Scalable serverless inference of AI models.
- Status of services**
Check if a specific AI4OS service might be down for some reason.
- Module template**
Create new modules based on our project's template.

You can also watch this [tutorial](#) for a quick overview of the project from the user's point of view (attached slides).



The Nomad cluster

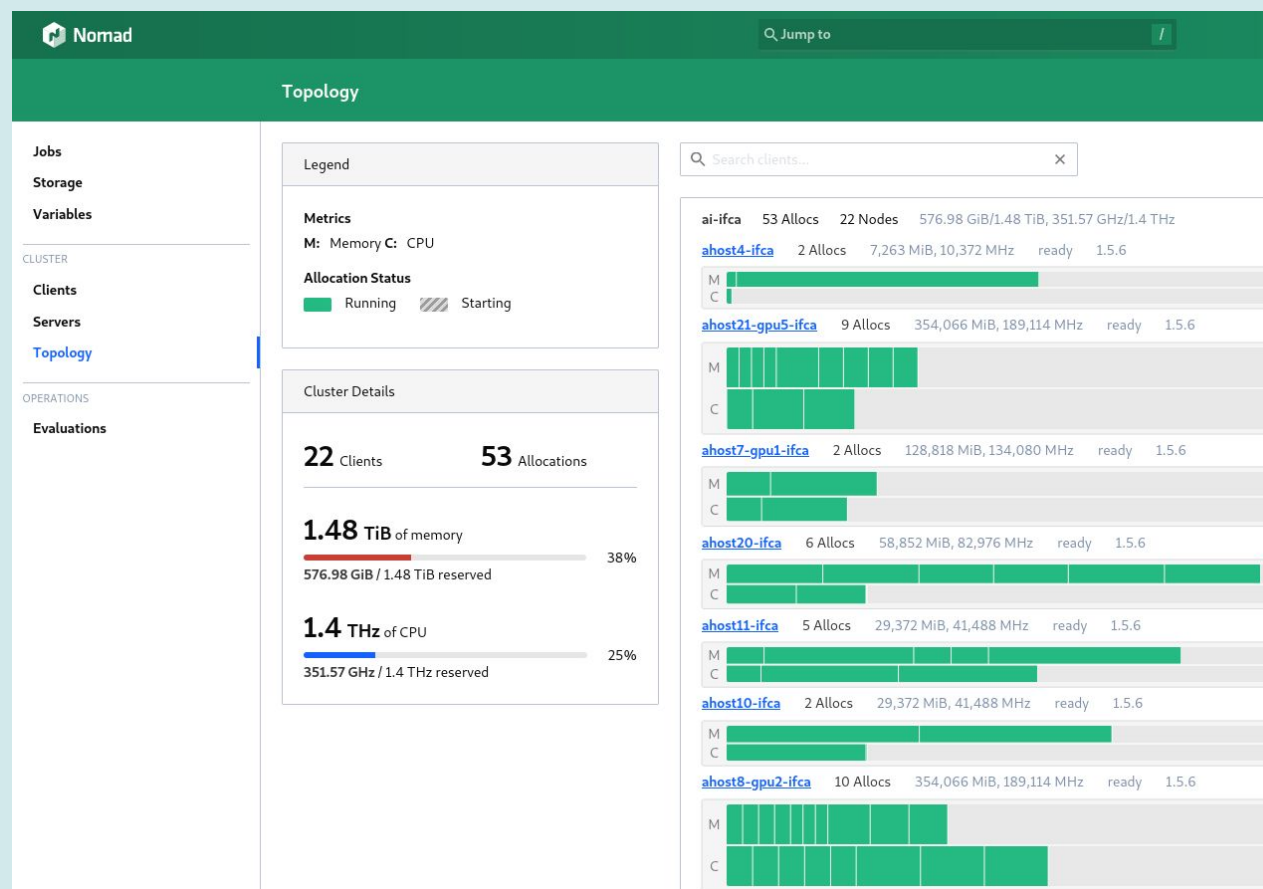
AI4

eosc



Co-funded by
the European Union

- one production cluster (IFCA) with 25 GPUs.
- multiple development clusters deployed for different tests (distributed/federated, full service mesh).
- deployments can make use of the “/storage” paths to automatically sync with Nextcloud (! slower access to files though)

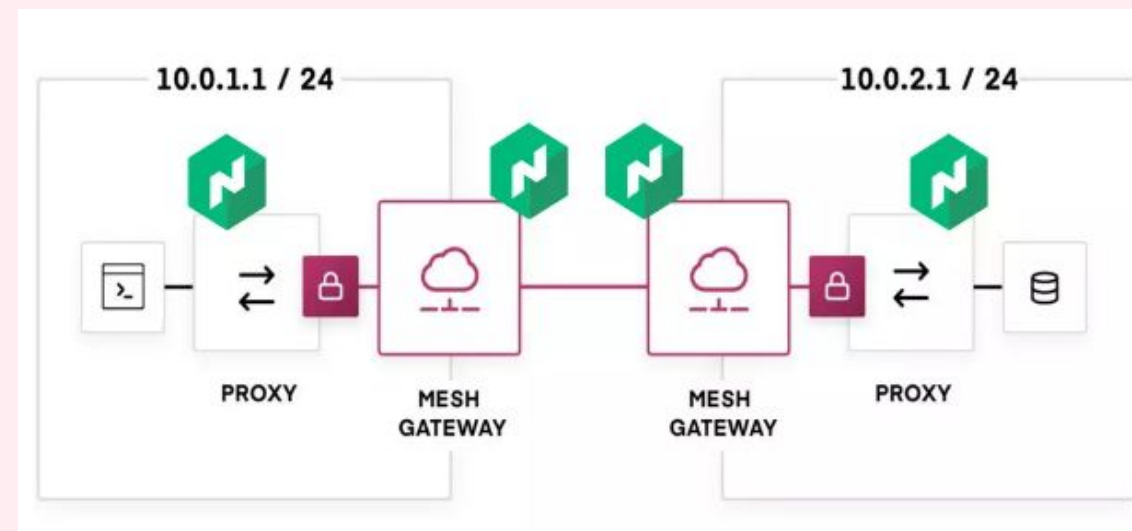
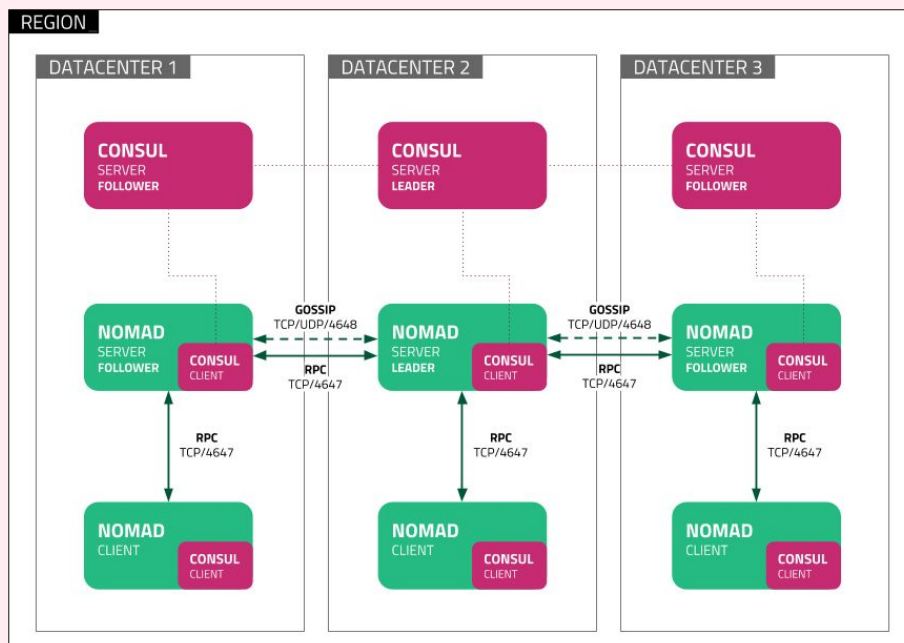




The Nomad cluster



- add more resources (GPUs, storage, ...)
- properly limit storage space inside each deployment
- improve data access (share volumes between deployments to avoid copying data from Nextcloud)
- federate multiple providers (IISAS, LIP, ...)





The Nomad cluster

AI4 |  eosc



Co-funded by
the European Union

 **November 27th** yearly maintenance shutdown

Please, don't forget to do a **backup** of your running jobs!

PAPI + Dashboard

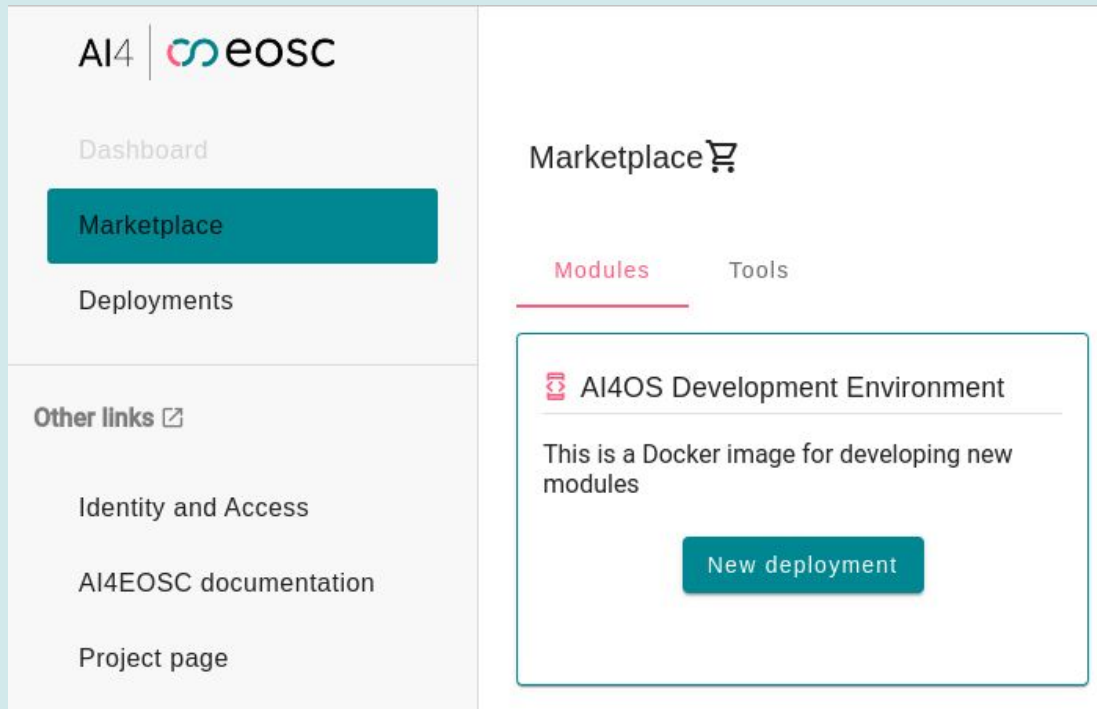
AI4

eosc

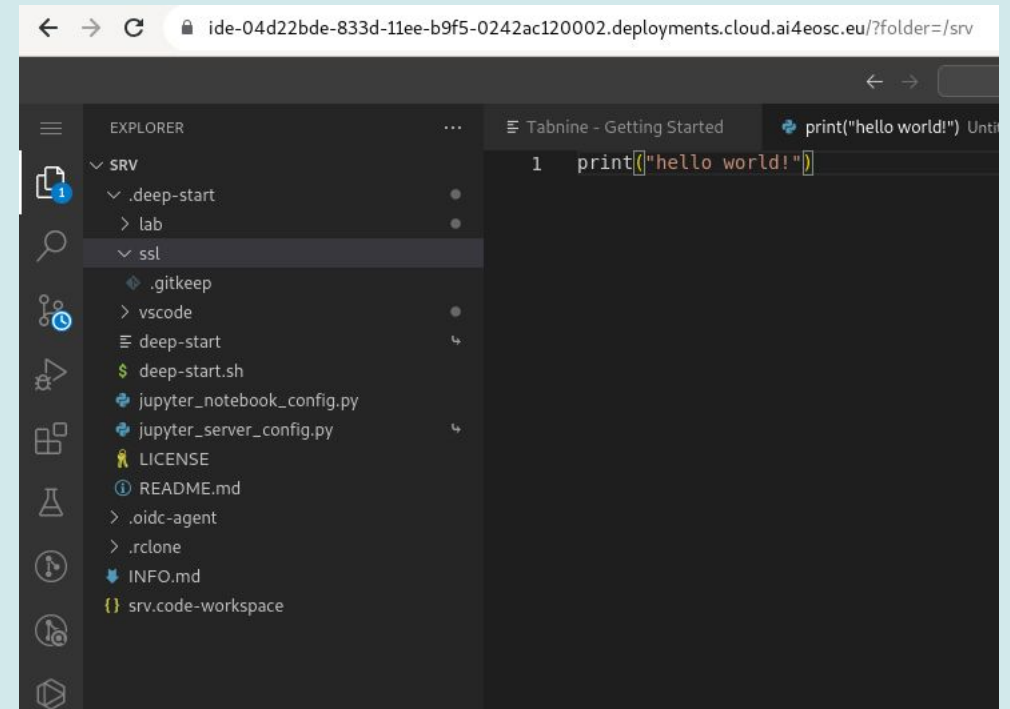


Co-funded by
the European Union

- Tools/Deployments view
- SSL in deployments
- Dev module with main DL frameworks and VS code installed
- Federated server tool (flower.io) available to perform federated trainings



The screenshot shows the AI4EOSC dashboard. On the left, there is a sidebar with the AI4EOSC logo and navigation options: Dashboard, Marketplace (highlighted in a teal box), and Deployments. Below this, there is a section for 'Other links' with options for Identity and Access, AI4EOSC documentation, and Project page. The main content area is titled 'Marketplace' and has two tabs: 'Modules' (selected) and 'Tools'. Under the 'Modules' tab, there is a card for 'AI4OS Development Environment' with a description: 'This is a Docker image for developing new modules' and a teal 'New deployment' button.



The screenshot shows a VS Code IDE interface. The browser address bar at the top indicates the URL: 'ide-04d22bde-833d-11ee-b9f5-0242ac120002.deployments.cloud.ai4eosc.eu/?folder=/srv'. The Explorer sidebar on the left shows a file tree for the 'srv' folder, including subfolders like '.deep-start', 'lab', and 'ssl'. The 'ssl' folder is expanded, showing files like '.gitkeep', 'vscode', 'deep-start', 'deep-start.sh', 'jupyter_notebook_config.py', 'jupyter_server_config.py', 'LICENSE', 'README.md', '.oidc-agent', '.rclone', 'INFO.md', and 'srv.code-workspace'. The main editor area shows a code editor with a single line of Python code: '1 print("hello world!")'.



PAPI + Dashboard

AI4

|  eosC



Co-funded by
the European Union

- Experiment centric view:
 - 1 experiment gathers several deployments and tools
 - ability to share experiments across users
- Visualize cluster stats in the Dashboard:
 - free resources
 - current user usage
 - historic user usage
 - historic VO usage
- Filter modules by tags
- New tools to be added:
 - MLflow (experiment tracking)
 - CVAT (image annotation)
- Deploy services specifically tailored for inference (ie. automatic scaling) (using OSCAR under the hood)
- Authentication for clients in federated trainings via token



Modules

AI4

|  eOSC



Co-funded by
the European Union

- Module creation with AI4OS template (cookiecutter)
- Module metadata in JSON
- DEEPaaS API to expose functionality of the module



Modules

- (?) Unify repos
- (?) Easier Jenkinsfiles
- Metadata:
 - move to YAML format
 - more tag categories
 - remove clutter
- DEEPaaS:
 - (?) type hints to define input args
 - (?) decorator use decorator instead of entrypoints
 - Integrate [Gradio UI](#) inside main DEEPaaS for inference endpoints

AI4

eosc



Co-funded by
the European Union

demo_app

A minimal toy application for demo and testing purposes. We just implemented dummy inference, ie. we return the same inputs we are feed.

DEMO-STR
some-string

DEMO-STR-CHOICE
choice2 ▾

DEMO-INT
1


DEMO-INT-RANGE
50

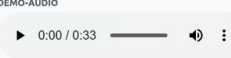
DEMO-FLOAT
0,1

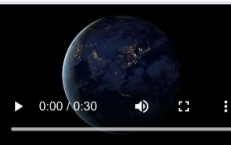
DEMO-BOOL

DEMO-DICT
{ "a": 0, "b": 1 }

DEMO-LIST-OF-FLOATS
0,1,0,2,0,3


DEMO-IMAGE
 Edit

DEMO-AUDIO


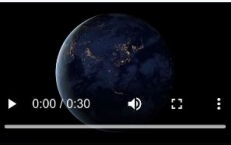
DEMO-VIDEO


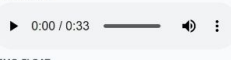
Clear Submit

DEMO-STR 0.73s
some-string

DEMO-IMAGE


DEMO-INT-RANGE
50

DEMO-VIDEO


DEMO-AUDIO


DEMO-FLOAT
0,1

DEMO-DICT
root: {} 2 keys
a: 0
b: 1

DEMO-BOOL
True

DEMO-STR-CHOICE
choice2

DEMO-INT
1

DEMO-LIST-OF-FLOATS
[0,1, 0,2, 0,3]

CLASSIFICATION SCORES

class3

class3	35%
class0	20%
class2	18%
class1	14%
class4	14%

Screenshot Flag



Others

- Improved CI/CD management (dedicated machine)
- Project auth:
 - Migration to EGI production auth (iMagine users)
 - (?) Migration to IAM (AI4EOSC users)
- Migration to new Nextcloud instance (hopefully transparent)
- Compose multistep inference flows with OSCAR, Elyra and Node-red.



What do you want to see in the platform?

feature requests - rough edges - enhancements - ...

AI4 |  eosc



Co-funded by
the European Union