

Open AiiDA lab platform for cloud computing in Materials Science

Principal Investigators: Giovanni Pizzi <giovanni.pizzi@epfl.ch>, Leopold Talirz <leopold.talirz@epfl.ch>

Shepherds: [Enol Fernandez del Castillo](#)

Entry in the community requirement database: [CR: Open AiiDALab platform for cloud computing in Materials Science \(EAP\)](#)

About the pilot

The AiiDA lab brings the AiiDA workflow manager for computational science (www.aiida.net) to the cloud. While domain experts can install AiiDA on their own hardware, the AiiDA lab web platform gives novice users access to their personal pre-configured AiiDA environment in the cloud. AiiDA is a workflow manager for computational science with a strong focus on provenance, performance and extensibility. When executing a workflow, AiiDA records the provenance calculations performed, codes used and data generated in a directed acyclic graph tailored to provide full reproducibility of any given result. The AiiDA engine relies on a message queue in order to support high-throughput use cases of up to 50k calculations per hour, and the relational database backend enables performant queries on graphs of tens of millions of nodes. AiiDA (TRL 7-8) is used in production for high-throughput calculations.

For the AiiDA lab (TRL 6-7, <http://materialscloud.org/aiidalab>), we are currently operating 4 instances, two on Openstack VMs and two on kubernetes. One of the kubernetes instances is deployed on limited EOSC-hub test resources in the CESNET Czech computing centre, and already uses EGI check-in for authentication. The AiiDA lab uses docker for user containers, and kubernetes for orchestration. Users get persistent home volumes for active use (no long-term storage component).

The pilot will expand the capacity of instance deployed at CESNET to:

- Provide an open AiiDA lab for researchers in Europe capable of supporting ~100 concurrent users
- Support of the order of ~1000 users in the system
- Test scalable kubernetes set-up so resources can be adjusted to the load as required

Description of supported work

- Provide access to Kubernetes managed infrastructure to support the deployment and operation of on open AiiDA lab instance
- Support the authentication and authorisation of users into AiiDA lab with EOSC users

Use Cases

- Automatic simulation of materials properties via turn-key workflow execution and monitoring
- Reproducible computational-science simulations using AiiDA to track the provenance
- HPC simulations on the cloud

Team

| Participant | Role | Name and Surname |
|-------------|-------------------|---|
| EPFL | PI | Giovanni Pizzi |
| EPFL | PI | Leopold Talirz |
| EGI.eu | Shepherd | Enol Fernandez del Castillo |
| CESNET | Technical support | Mirek Ruda |
| | | |
| | | |

Technical Plan

The full technical plan can be found here:

| | |
|------------------|--|
| Preparation work | Integration with EGI Check-in demo instance Set up the instance with support for AiiDA 1.1 (released in Feb 2020) |
|------------------|--|

| | |
|------------------------------------|--|
| Work planned for Q2 (Apr-May-Jun) | <p>Increase capacity at CESNET to support at least 50 users (add: 50 CPUS, 150GB RAM, 1TB storage), setup k8s cluster with EC3</p> <p>Move to EGI Check-in production instance</p> <p>Deployment of an application to run Quantum ESPRESSO relaxation workflows in a fully automated fashion</p> |
| Work planned for Q3 (Jul-Aug-Sept) | <p>On-boarding of users (~50 expected)</p> <p>Installation in the instance of updated apps, software stack and data as needed for the tutorial</p> <p>Use of the AiiDA lab instance for a classroom/tutorial supporting at least 50 concurrent users (e.g. for the tutorial we'll hold in Lithuania in July 2020: https://www.cecam.org/index.php/workshop-details/3 or a later one)</p> |
| Work planned for Q4 (Oct-Nov-Dec) | <p>On-boarding of users (>100 expected), and testing scaling to at least a few hundreds. Identification of issues (e.g. required disk /CPU/RAM quotas, speed and efficiency of the storage to support efficient database storage and querying)</p> <p>Upgrade of the full software stack to the most recent available software (AiiDA version, AiiDA plugins, python libraries).</p> <p>Provide straightforward apps to connect to HPC resources (for which the user has already credentials)</p> <p>Support of at least 6 different simulation codes and the corresponding AiiDA plugins (e.g. among these: Quantum ESPRESSO, Siesta, Fleur, BigDFT, Yambo, Wannier90, CP2K) (to be clarified)</p> |

EOSC services and providers

Providers

- EGI Check-in
- EGI Cloud Container Compute and EGI Online Storage (CESNET)
- IM/EC3 (UPV)

Services

- EOSC-hub AAI (Check-in): support for integration, simplification of the registration process
- EGI Cloud Container Compute (CESNET):
 - Support for deployment of scalable Kubernetes
 - Provisioning of resources
- EGI Online Storage (CESNET):
 - 10TB storage available as Kubernetes volumes
- IM/EC3:
 - Auto-scaling Kubernetes setup
- B2HANDLE/Zenodo:
 - Explore PID releasing services/Zenodo integration for the Materials Cloud Archive
- EOSC-hub monitoring:
 - A/R monitoring of the service