Designing and Enabling E-infrastructures for intensive Processing in a Hybrid DataCloud

HORIZON 2020

Designing and Enabling E-infrastructures for intensive Processing in a Hybrid DataCloud

Rapports

Informations projet Financé au titre de **DEEP-HybridDataCloud EXCELLENT SCIENCE - Research Infrastructures** N° de convention de subvention: 777435 Coût total € 2 988 750,00 Site Web du projet 🔼 Contribution de DOI ľUE 10.3030/777435 🔼 € 2 988 750,00 Coordonné par Projet clôturé AGENCIA ESTATAL CONSEJO SUPERIOR DE Date de signature de la CE INVESTIGACIONES CIENTIFICAS 31 Octobre 2017 Spain Date de début Date de fin 1 Novembre 2017 30 Avril 2020

Periodic Reporting for period 2 - DEEP-HybridDataCloud (Designing and Enabling E-infrastructures for intensive Processing in a Hybrid DataCloud)

Période du rapport: 2019-02-01 au 2020-04-30

Résumé du contexte et des objectifs généraux du projet

"The DEEP-HybridDataCloud project supports intensive computing techniques (in particular machine learning, deep learning and artificial intelligence) that require the usage of specialized HPC hardware, like GPUs or low-latency interconnects, to explore very large datasets. Leveraging on mature technologies from previous EU actions (like INDIGO-DataCloud) and external open source projects the project integrates all the components required to provide specialized and high level added value services focused on machine learning, deep learning and artificial intelligence over distributed pan-European e-Infrastructures, following a hybrid-cloud approach. Therefore, DEEP-HybridDataCloudd provides a comprehensive framework that streamlines the development of these applications, covering the whole application life-cycle: from the model creation, update, training and evaluation, publication and reuse.

The DEEP platform (<u>https://deep-hybrid-datacloud.eu/the-platform/</u>) is composed of several specialized high level services that can be used individually or as a whole. The "DEEP training facility"" (https://train.deep-hybrid-datacloud.eu/ 1) allows data scientists and machine learning practitioners to develop and train their applications exploiting existing specialized hardware in EU e-Infrastructures. This service allows to either create their own model using our development environment or to reuse an existing one. To this aim, the "DEEP marketplace"" (https://marketplace.deep-hybrid-datacloud.eu/ 1/2) allows to publish, share and reuse machine learning, deep learning and artificial intelligence models easily, creating a knowledge hub for reuse and collaboration, bringing knowledge closer to society. Lastly, the ""DEEP as a Service"" (https://deepaas.deep-hybrid-datacloud.eu/ 2) provides an automated way to transparently publish an existing model as a service, with horizontal scalability out of the box. The project has followed a Service Oriented Architecture (SOA), with an emphasis on the quality assurance of the developed tools and applications, than undergo a comprehensive testing and validation process before being released or deployed. Moreover, all the developed applications can be transparently executed on a disparate set of resources, ranging from personal laptops, workstations, EU e-Infrastructures or HPC facilities just to cite some of them.

These services are integrated with storage and identity services in the European Open Science Cloud, and are offered through the EOSC portal, providing therefore a framework for machine learning, deep learning and artificial intelligence for the EOSC, exploiting first-class EU e-Infrastructure resources."

Travail effectué depuis le début du projet jusqu'à la fin de la période considérée dans le rapport et principaux résultats atteints jusqu'à présent

"The first half of the project resulted in the launch of the first platform and testbed prototype (codenamed DEEP-Genesis). After this initial prototipe was launched, an extensive testing phase was performed, together with use cases broth within the project as well as external communities, in order to assess and validate the platform functionality. This has led to the 2nd project release, producing the DEEP-Rosetta platform. DEEP-Rosetta is composed of three different services, as depicted in the

attached diagram, resulting in the service offer included in the DEEP-HybridDataCloud service catalog towards the EOSC. These services can be used jointly as a whole framework, or individually and are the following:

• The DEEP training facility (<u>https://train.deep-hybrid-datacloud.eu/</u> as the service providing access to resources leveraging accelerators for intensive computing, exploiting the PaaS cloud layer or HPC resources.

• The DEEP as a Service (<u>https://deepaas.deep-hybrid-datacloud.eu/</u>], which provides an easy and automated way to deploy existing applications in the catalog as services.

• The DEEP Open catalog (<u>https://marketplace.deep-hybrid-datacloud.eu/</u>) a marketplace where users can browse, retrieve, share and store their applications.

• The data and storage services, where user data and results can be stored that are external to the project, but integrated with us.

These services are being deployed on resource providers that are part of production pan-European e-Infrastructures. In this regard we have ensured that the deployed services are in line with the operational requirements that are enforced at the e-Infrastructure level. Besides, we have linked with industry stakeholders through the EOSC Digital Innovation Hub (DIH).

Extensive work has been performed in all the software quality and maintenance related activities, involving both software developers and users since the initial phases. This co-design methodology was carried out in order to obtain a high degree of acceptance on the software related procedures, avoiding excessive burden on software development teams. Moreover, these activities are being done in coordination with other projects, as reflected in the publication of ""A set of common software quality assurance baseline criteria for research projects" document (<u>http://hdl.handle.net/10261/160086</u>). Besides, we have ensure that the developed services follow the existing recommendations to be integrated into the EOSC realm."

Progrès au-delà de l'état des connaissances et impact potentiel prévu (y compris l'impact socio-économique et les conséquences sociétales plus larges du projet jusqu'à présent)

DEEP-HybidDataCloud provides a comprehensive framework for the development of machine learning, deep learning and artificial intelligence applications for the European Open Science Cloud. We build on top of pan-European e-Infrastructures, following a Service Oriented Architecture, in order to deliver high level and added value services for scientists and the public in general. We provide transparent access to e-Infrastructures both for training and deploying machine learning, deep learning and artificial intelligence models, covering the whole machine learning life cycle. Our framework and reference implementation is based on an approach where we differentiate our users by their knowledge in different areas (domain knowledge, machine learning knowledge, technical knowledge). Therefore we offer them a path where it is possible to perform different tasks, according to their needs and their prior skills. Advanced users will get the more advanced features, and basic users will still obtain enough functionality form the framework. In this regard, the DEEP framework has

been proposed as the reference specification for the machine learning, deep learning and data analytics technical specification for the EOSC Workflow management and user interfaces and Data analytics Working Group.

We bring knowledge closer to society by easing the execution of the DEEP marketplace modules. We provide easy methods (i.e. easy to users with no technical knowledge) and instructions to quickly test and use the model's functionality and exploit its gained knowledge. Moreover, we allow scientists to deploy their applications and services, being possible therefore to build services that exploit the model's functionality. We are fostering collaboration since all the developed modules, applications and models are directly available for download and reuse. Reproducibility, although not being directly addressed by our project, can be achieved by the publication of a model, application, data and metadata through the portal.



1st release banner



DEEP-HybridDataCloud logo

Dernière mise à jour: 24 Septembre 2020

Permalink: https://cordis.europa.eu/project/id/777435/reporting/fr

European Union, 2025