

Machine learning and deep learning services for the EOSC

The DEEP-Hybrid-DataCloud project

EGI Conference 2020

November 2 - 5, 2020

Álvaro López García aloga@ifca.unican.es Spanish National Research Council







DEEP project in 1 slide



- Designing and Enabling E-Infrastructures for intensive data Processing in a Hybrid DataCloud (Grant agreement number 777435, Nov 2017 - Apr 2020)
- Global objective: Promote the use of intensive computing services by different research communities and areas, and the support by the corresponding e-**Infrastructure** providers and open source projects
 - Focusing on Machine learning, Deep learning, and Post processing services for the European Open Science Cloud





















HelmholtzZentrum münchen

Deutsches Forschungszentrum für Gesundheit und Umwelt

Context: the EOSC



- The European Open Science Cloud was launched in 2016 as part of the European Cloud Initiative
- The EOSC is a digital platform to support and develop science and open innovation
- Europe's virtual environment for all researchers
- Providing seamless access to data and interoperable services across borders

- Access is delivered through the EOSC portal
- Launched in 2018, continuously being enhanced and developed



DEEP-HybridDataCloud vision

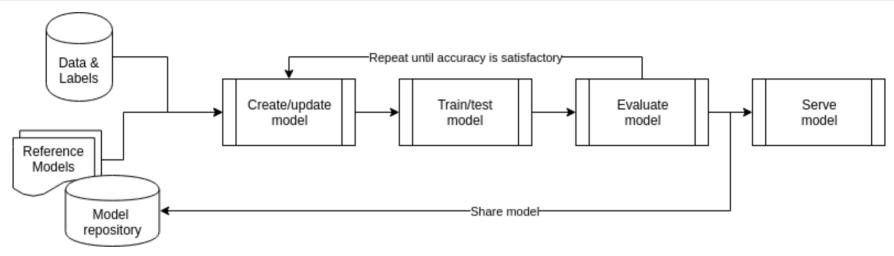


- Provide specialized cloud services to develop, exploit and share machine learning and deep learning applications → Service-Oriented Architectures and platforms
 - Covering the whole machine learning application development cycle
 - Focused on all types of users and user knowledge
- Transparent access to specialized computing resources (accelerators, high performance computing) → reduce entry barrier
- Build an EOSC machine learning marketplace as an application exchange → ease of use, foster collaboration, knowledge dissemination
- Run the same application everywhere (latptop, cloud, HPC, etc.) → ease of use



Machine learning development cycle





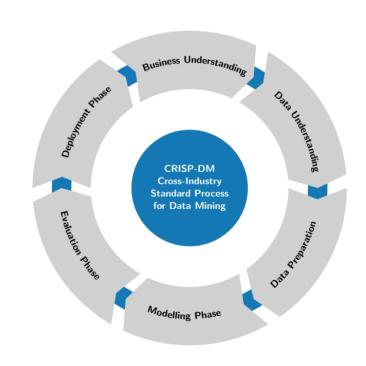
- The project covers all development phases of a machine learning application:
 - Creation, update, improvement of a model (existing or new)
 - Model training, test, evaluation
 - Model deployment as a service
 - Model publication for sharing and reuse

5/30

The CRISP-DM cycle



- We are focusing on several phases of the Cross-Industry Standard Process for Data Mining (CRISP-DM) cycle:
 - Modeling, evaluation and deployment phases
- Provide a training and evaluation platform
- Provide a model deployment service
- Provide a catalog to share applications and models



Different users, different needs



- Category 1: Deploy a readily trained network for somebody else to use it on his/her data set
 - Domain knowledge
- Category 2: Retrain (parts of) a trained network to make use of its inherent knowledge and to solve a new learning task
 - Domain + machine learning knowledge
- Category 3: Completely work through the deep learning cycle with data selection, model architecture, training and testing
 - Domain + machine + technological knowledge

Docker containers



 DEEP leverages Docker containers to encapsulate the user application, together with the required environment



Containers are automatically built and published to the marketplace

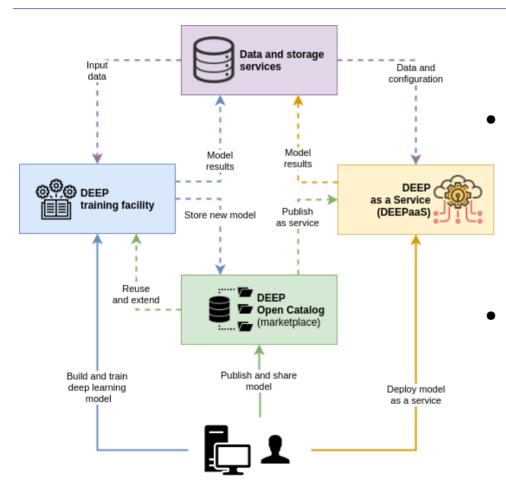


- Docker allows users to run the same application everywhere:
 - Using Docker, if available
 - Using uDocker (user space tool, no system-wide install needed) if Docker is not available (like in HPC environments)
 - See: https://github.com/indigo-dc/udocker/
 - Using serverless components (like Apache OpenWhisk)



The DEEP services





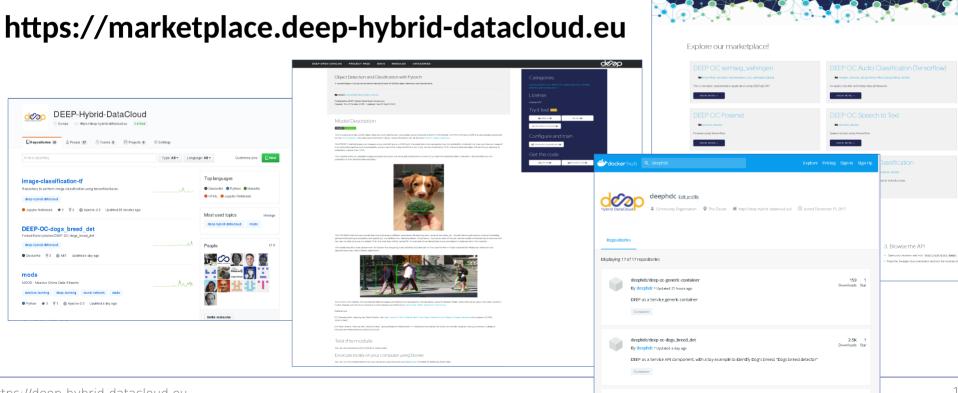
- High level specialized services for exploitation through EOSC
 - DEEP Open Catalog → sharing
 - DEEP training facility → training
 - DEEP as a Service → deployment
- Integration with storage from external initiatives (e.g. eXtreme-DataCloud, EGI Data Hub, etc.)

https://deep-hybrid-datacloud.eu 10/30

DEEP Open Catalog

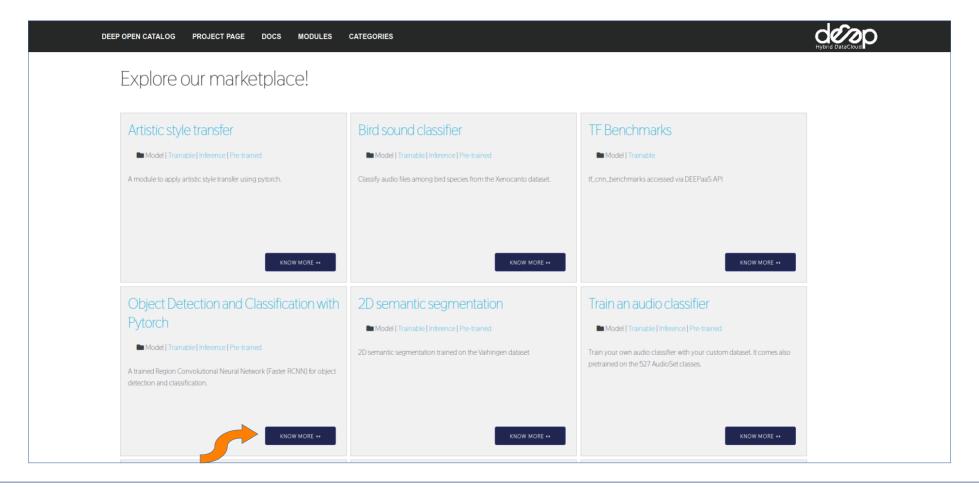


Sharing knowledge, models and applications between users



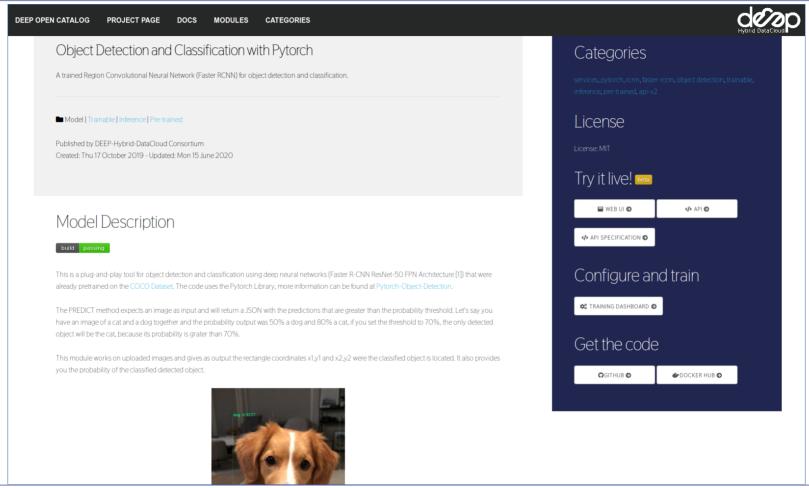
Browsing the marketplace





Marketplace entry





https://deep-hybrid-datacloud.eu 14/30

Local inference and prediction



```
(udocker) alvaro:~ $ udocker pull deephdc/deep-oc-obj_detect_pytorch
Info: downloading layer sha256:2746a4a261c9e18bfd7ff0429c18fd7522acc14fa4c7ec8ab37ba5ebaadbc584
(wdocker) alvaro:~ $ udocker create deephdc/deep-oc-obj_detect_pytorch
(udocker) alvaro:~ $ udocker run -p 5000:5000 deephdc/deep-oc-obj_detect_putorch
Warning: this container exposes TCP/IP ports
Warning: non-existing user will be created
                STARTING 9bb8179d-c731-348c-bbf4-e43045ee8936
 executing: deepaas-run
             Hybrid-DataCloud ##
Welcome to the DEEPaaS API API endpoint. You can directly browse to the
API documentation endpoint to check the API using the builtin Swagger UI
or you can use any of our endpoints.
   API documentation: http://0.0.0.0:5000/ui
   API specification: http://0.0.0.0:5000/swagger.json
          V2 endpoint: http://0.0.0.0:5000/v2
```

Local inference and prediction

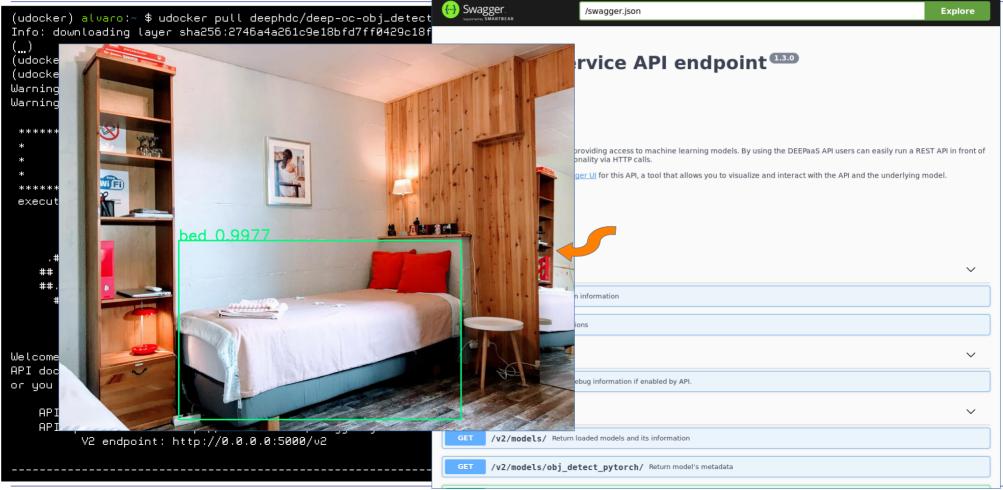


```
/swagger.json
(udocker) alvaro:~ $ udocker pull deephdc/deep-oc-obj_detect
Info: downloading layer sha256:2746a4a261c9e18bfd7ff0429c18f
(udocker) alvaro:~ $ udocker create deephdc/deep-oc-obj_dete
                                                                              DEEP as a Service API endpoint 1330
(udocker) alvaro:~ $ udocker run -p 5000:5000 | deephdc/deep-
Warning: this container exposes TCP/IP ports
Warning: non-existing user will be created
                                                                              This is a REST API that is focused on providing access to machine learning models. By using the DEEPaaS API users can easily run a REST API in front of
                     STARTING 9bb8179d-c731-348c-bbf4-e43045ee89
                                                                              their model, thus accessing its functionality via HTTP calls.
                                                                              Currently you are browsing the Swagger UI for this API, a tool that allows you to visualize and interact with the API and the underlying model.
                                                                                 · Project website.
 executing: deepaas-run
                                                                                 · Project documentation.
                                                                                 · Model marketplace
                                                                              API documentation
           ##
                                                                               versions
                                                                                       /v2/ Get V2 API version information
                 Hubrid-DataCloud
                                                                                         Get available API versions
                                                                               debug
Welcome to the DEEPaaS API API endpoint. You can directly br
API documentation endpoint to check the API using the builti
                                                                                       /v2/debug/ Return debug information if enabled by API.
or you can use any of our endpoints.
                                                                               models
    API documentation: http://0.0.0.0:5000/ui
    API specification: http://0.0.0.0:5000/swagger.json
                                                                                       /v2/models/ Return loaded models and its information
            V2 endpoint: http://0.0.0.0:5000/v2
                                                                                       /v2/models/obj detect pytorch/ Return model's metadata
```

https://deep-hybrid-datacloud.eu

Local inference and prediction





https://deep-hybrid-datacloud.eu 17/30

DEEPaaS: DEEP as a Service

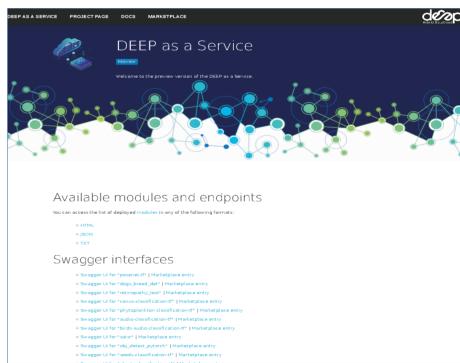


Automatic deployment of catalog modules to exploit their functionality

https://deepaas.deep-hybrid-datacloud.eu







https://deep-hybrid-datacloud.eu 18/30

DEEPaaS: automated deployment





Available modules and endpoints

You can access the list of deployed modules in any of the following formats:

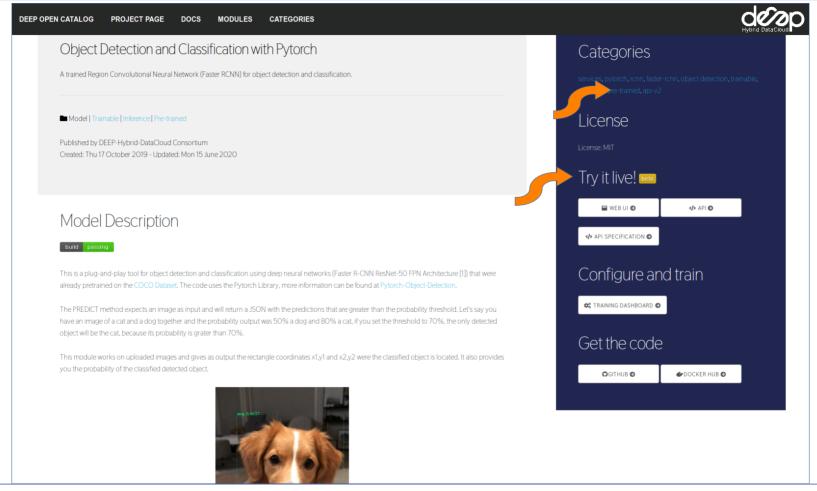
- HTML
- JSON
- TXT

Swagger interfaces

- o Swagger UI for "image-classification-tf-dicom" | Marketplace entry
- Swagger UI for "dogs breed det" | Marketplace entry
- Swagger UI for "semseg vaihingen" | Marketplace entry
- Swagger III for "obj. detect. pytorch" I Marketplace entre

Marketplace entry

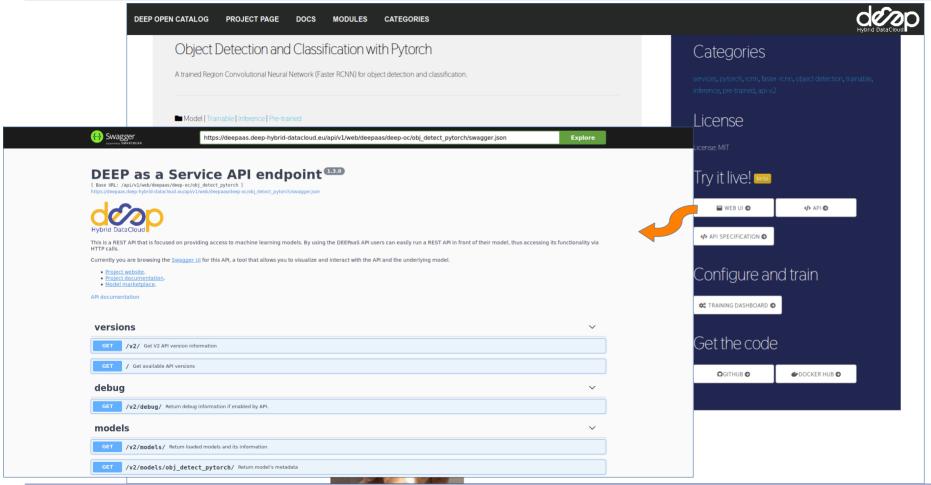




https://deep-hybrid-datacloud.eu 20/30

Marketplace entry





https://deep-hybrid-datacloud.eu 21/30

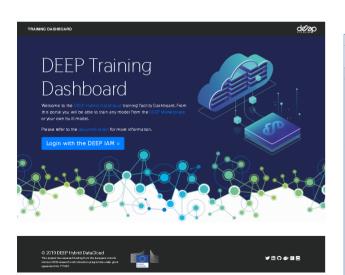
DEEP training facility

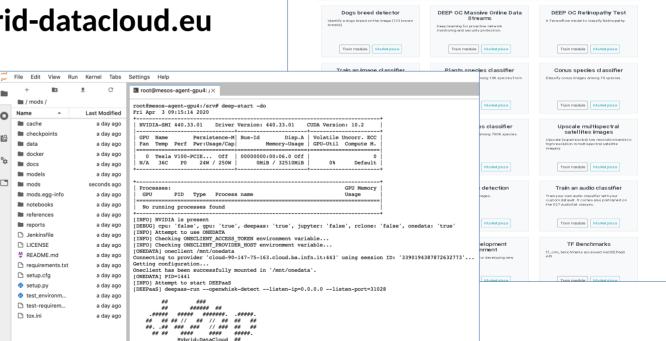


€ ALVARO LOPEZ GARCIA -

1-click development and training environments in Cloud and HPC resources

https://train.deep-hybrid-datacloud.eu





DEPLOYMENTS

Explore the Marketplace

A DVA NCED

avtancibla mannar

EXTERNAL LINKS

DEEP Development Environment

The DEEP Development Environment provides a ready to

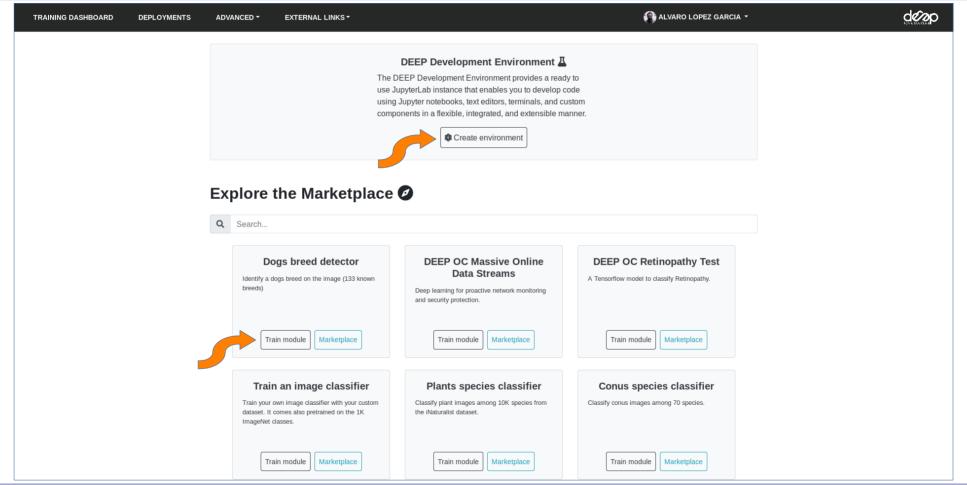
de Create environment

use JupyterLab instance that enables you to develop code using Jupyter notebooks, text editors, terminals, and custom components in a flexible, integrated, and

https://deep-hybrid-datacloud.eu 23/30

DEEP training dashboard

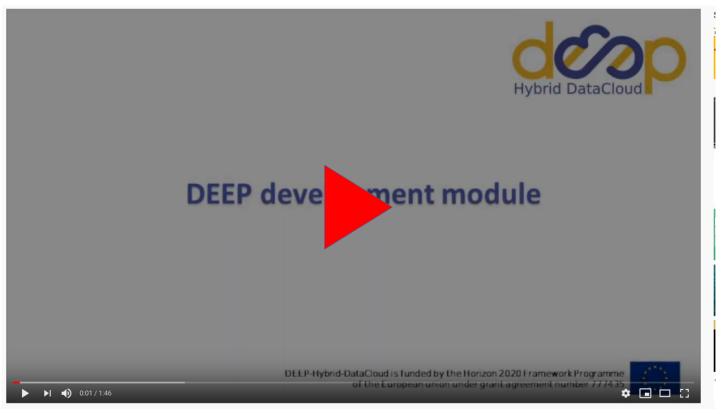




https://deep-hybrid-datacloud.eu 24/30

DEEP development environment



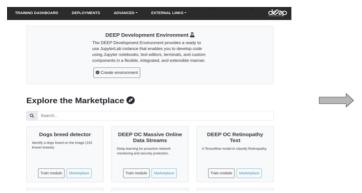


https://youtu.be/J_I_xWiBGNA

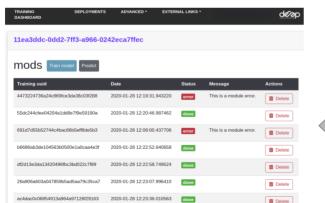
DEEP training facility



1. Select a module from the Open Catalogue



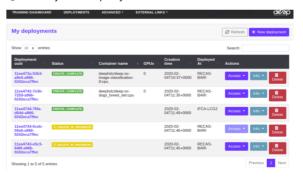
4. Monitor the training history



2. Choose your configuration parameters



3. View your deployments

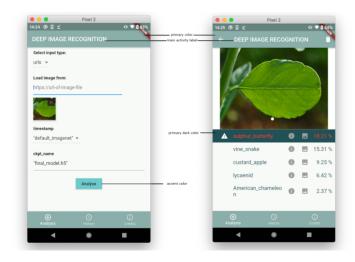


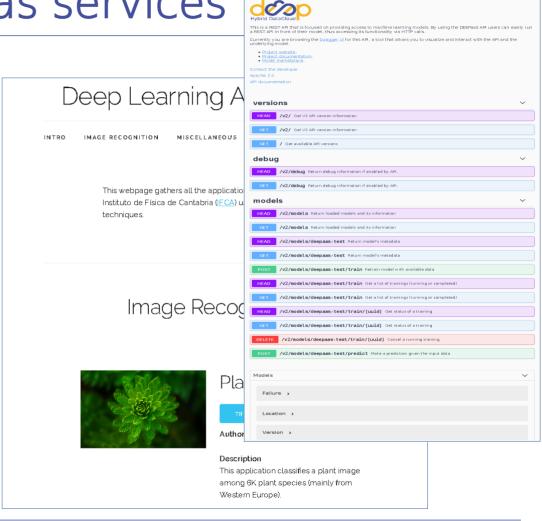
27/30

DEEPaaS API: models as services

Exposing models through a common and standard-based API

https://deepaas.readthedocs.io



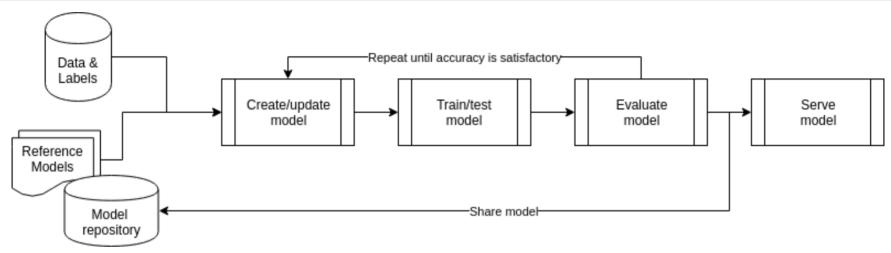


DEEP as a Service API endpoint (100.0dev32)

https://deep-hybrid-datacloud.eu 28/30

Machine learning development cycle





- The project covers all development phases of a machine learning application:
 - Creation, update, improvement of a model (existing or new) → DEEP Open Catalog and DEEP training facility
 - Model training, test, evaluation → DEEP training facility
 - Model deployment as a service → DEEP as a Service
 - Model publication for sharing and reuse → DEEP Open Catalog

30/30

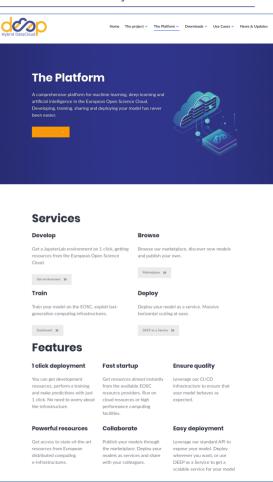
DEEP-Rosetta: 2nd platform and release





Second software release and platform released January 2020

- Full support for the whole machine learning application life cycle
 - Sharing and reuse through DEEP Open Catalogue
 - Training, testing and evaluation on DEEP training facility
 - Live deployment on DEEPaaS
- Transparent access to Cloud and HPC resources



https://deep-hybrid-datacloud.eu 31/30

Conclusions



- DEEP provides a comprehensive, service-oriented, framework to develop machine learning, deep learning and artificial intelligence applications
 - Exploiting distributed e-Infrastructure and HPC resources, with special focus on the EOSC
 - Covering the whole development phases (CRISP-DM and machine learning lifecycle)
 - High level and added value services that can be used standalone or as a whole platform solution
- Services at production level with TRL>=8
- Industry opportunities: engagement of industrial stakeholders through collaboration with EOSC-Digital Innovation Hub

PECIAL SECTION ON SCALABLE DEEP LEARNING FOR BIG DATA

EEE Access

Received December 20, 2019, accepted December 31, 2019, date of publication January 6, 2020, date of current version January 30, 2020.

A Cloud-Based Framework for Machine Learning Workloads and Applications

ÁLVARO LÓPEZ CARCÍA[®] J. JESÚS MARCO DE LUCA[®] J. MARICA ANTONACCI[®] J. WOLFGANG ZU CASTELL[®] 3.19, MARIO DAVID[®] J. MARCUS HARDT[®] S. LARA LLORET [GLESIAS[®] J. GERMÁN MOLTÓ [®] J. MARCUS PLOCIENNIK[®] J. VIET TRAN[®] S. ANDY S. ALIC[®] J. MICUEL CABALLER [®] J. ISABEL CAMPOS PLASENCIA[®] J. ALESSANDRO COSTANTINI[®] J. STEFAN DILOGINISKY[®] 9. DOINA CRISTINA DUMA[®] GIACINTO DONVITO[®] S. JORGE COMES[®] J. IGNACIO HEREDIA CACHA[®] J. KEIICHI ITO[®] J. VALENTIN Y. KOZLOW[®] GIANG NGUYEN[®] J. PABLO ORVIZ FERNÁNDEZ[®] J. ZOBEŘ ŠÚSTÉ[®] J. AND PAWEL WOLNIEWICZ[®] J.

IFCA (CSIC-UC), 39005 Santander, Spain
²Laboratory of Instrumentation and Experimental Particle Physics, 1600-078 Lisbon, Portuga

*Lanoratory of instrumentation and experi TNFN Bari, 70126 Bari, Italy *INFN CNAF, 40127 Bologna, Italy

FOrman Supercomputing and Networking Center, IBCh PAS, 61-704 Pozman, Poland

6 Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany

Instituto de Instrumentación para Imagen Molecular (I3M), CSIC, Universitat Politêcnica de Valência, 46022 Valencia, Spain

Insulus de inseruncimento qua integra sociedar (1931, 2512, de desenta i rotacida de varieta de varieta de l'anticida el finalment el f

¹⁰De partment of Mathematics, Technische Universität München, 85748 Munich, Germany LICESPET, 16000 Berner, Corole Berneley

Corresponding author: Álvaro López García (aloga@ifca.unican.es

This work was supported by the project DEIP-Hybrid-DataCloud "Designing and Enabling E-infrastructures for intensive Processing in a Hybrid DataCloud" that has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Gantt 771435.

ABSTRACT In this paper we propose a distributed architecture to provide machine learning practitioners with a set of tools and cloud services that cover the whole machine learning development cycle: ranging from the models creation, training, validation and testing to the models serving as a service, sharing and publication. In such respect, the DEEP-Hybrid-DataCloud framework allows transparent access to existing clinfarstructures, effectively exploiting distributed resources for the most computer intensive tasks coming from the machine learning development cycle. Moreover, it provides scientists with a set of Cloud-oriented services to make their models publicly available, by adopting a serverless architecture and a DevOps approach, allowing an easy share, publish and deploy of the developed models.

INDEX TERMS Cloud computing, computers and information processing, deep learning, distributed computing, machine learning, serverless architectures.

INTRODUCTION

The impact of emerging computing techniques together with an increasing dimension of large datasets and the availability of more and more performing and accessible computing resources is transforming many research areas. This opers the door to new opportunities to tackle unprecedented research challenges. Over the last decade there has been a boost on the usage of machine learning techniques in most of the research

The associate editor coordinating the review of this manuscript and approving it for publication was Liangxiu Han.

areas, and recently it has even improved with the adoption of deep learning techniques, e.g. LeCun et al. [1]. Although the basic components of the techniques are well known, recent advances arouse the interest from the scientific community towards this area, and it has already become a state-of-the-art technology in many fields, from computer vision to speech

The performance increase in the existing computing technologies and the availability of specialized computing devices played an important role in the advent of deep learning. In particular, the availability of more efficient and

VOLUME 8, 2020 This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

1868

Links



Request access through EOSC portal

https://marketplace.eosc-portal.eu/services/deepaas-training-facility

DEEP Services

Marketplace: https://marketplace.deep-hybrid-datacloud.eu

Training facility: https://train.deep-hybrid-datacloud.eu/

DEEPaaS: https://deepaas.deep-hybrid-datacloud.eu/

Documentation

http://docs.deep-hybrid-datacloud.eu

DEEP videos

https://www.youtube.com/playlist?list=PLJ9x9Zk1O-J_UZfNO2uWp2pFMmbwLvzXa

Contact



Contact

aloga@ifca.unican.es, deep-info@listas.csic.es

Web page

https://deep-hybrid-datacloud.eu/the-platform/

Social media

https://twitter.com/DEEP_eu

https://twitter.com/IFCA_Computing

Thank you Any Questions?



