DEEP-HybridDataCloud

G. Donvito (INFN), J. Gomes (LIP), A. Juan Ferrer (Atos), V. Kozlov (KIT),
A. López García (CSIC), L. Matyska (CESNET), N. Meyer (PSNC),
G. Moltó (UPV), V. Tran (IISAS), W. zu Castell (HMGU)

deep Hybrid DataCloud
https://deep-hybrid-datacloud.eu

Goal: prepare **a new generation of e-Infrastructures** that harness latest generation technologies, supporting **deep learning** and other **intensive computing** techniques to exploit **very large data** sources

Objective: promote the use of intensive computing services by different research communities and areas, and their support by the corresponding e-Infrastructure providers and open source projects

Keywords: *Clouds and Distributed Computing, Containerized HPC, Deep Learning, Accelerators, HPC Workflows*

Technologies to consider

Explore the applicability of **containers**

docker, nvidia-docker, UDOCKER, LXC / LXD ...

Either in **Bare-metal-like** (batch systems)
or in **Cloud middleware**

openstack (Magnum, nova-lxd), Open Nebula (ONEDock), kubernetes, Apache MESOS (Marathon, Chronos)

Key challenges

- Scalability for containerised HPC Applications
- Native support for GPU and Infiniband
- Enable containers for unprivileged users

Solution strategies

- Testbeds with GPU and Infinband
- Expand computing capability to address multiple resource providers
- Use containers to fully utilise bare-metal resources in the Cloud
- Offer a DevOps approach for the development of applications
- Analysis and implementation of pilot applications

Roadmap

- Final report after testing at large scale
- Final release of High Level Hybrid Cloud solutions deployed
  Application-as-a-service implementation
  Second prototype of the Pilot testbed and integration with EOSC infrastructures, supporting all use-cases
  Functionalities and scalability of High Level Hybrid Cloud solutions demonstrated
  DEEP-as-a-Service solution

2020

- Final implementation for supporting accelerators and HPC
- Use-cases are integrated in testbeds, DevOps approach operational

2019

- First prototype of the Pilot testbed and integration with EOSC, supporting some use-cases
  First implementation of the software platform
  High Level Hybrid Cloud solutions prototype completed
  Batch executions in the DEEP-as-a-Service solution
- Technology assessment and initial selection of use-cases
  Initial design and implementation plan for software platform
  High Level Hybrid Cloud solutions general architecture defined
  DEEP-as-a-Service architecture and API description

2018

- Assessment of available technologies carried out

Start

Use-cases

Biology

- Deep learning for retinopathy detection
  automated classification and stage and progression of retinopathy based on large set of color fundus retinal photography images [1]
- Plant Classification with Deep Learning
  automatically identify plant species from images using deep learning [2]

Earth observation

- Deep learning application for monitoring through satellite imagery
  demonstrate the potential of combining satellite imagery and machine learning techniques in a cloud infrastructure [3]

Network Security

- Massive Online DataStreams (MODS)
  address security issues, intrusion detection, and anomaly detection by processing the information collected at the data centre level and using machine learning / deep learning techniques [4]

Physics

- Post-processing of massive amounts of data
  use cloud-based computing facilities and services to facilitate the analysis of Lattice QCD configurations in the course of the corresponding simulation [5]

References

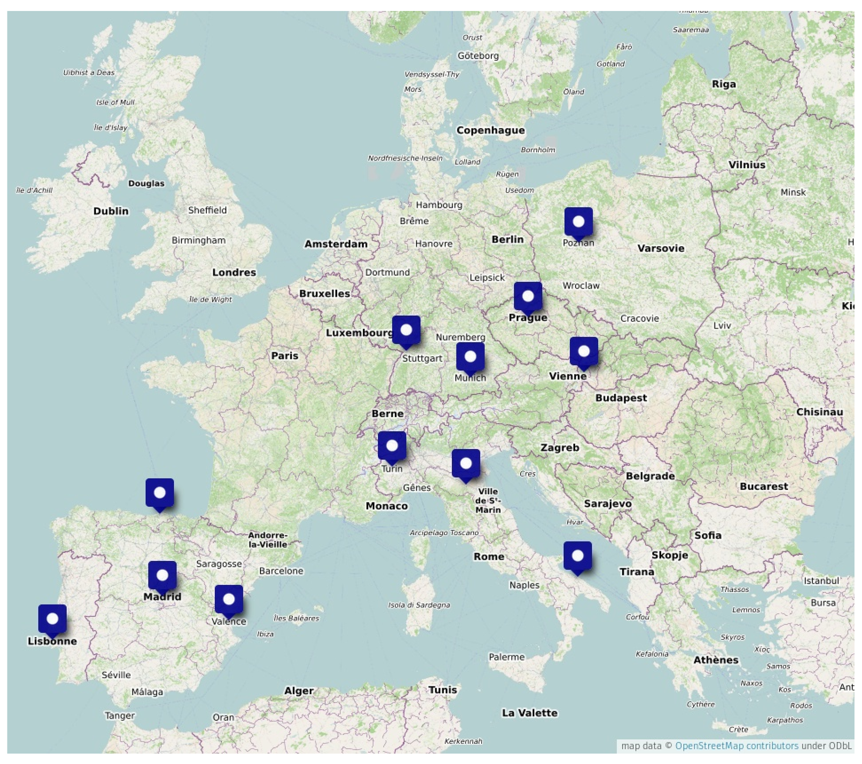
[1] Eulenberg, P. et al, Nature communications 8:463 (2017).
[2] Heredia, I., Proc. of the Computing Frontiers Conference (2017), p.259
[3] Jean, N. et al, Science (2016), Vol. 353, Issue 6301, p.790
[4] Nguyen, G., Data & Knowledge Engineering (2018), in press
[5] Campos, I., EPJ Web of Conferences (2018) 175, 09005

Key figures

Runtime: 30 months, November 2017 - April 2020

Partners: 9 academic and 1 industrial partner from 7 countries

Atos, cesnet, CSIC, INFN, HelmholtzZentrum münchen (Deutsches Forschungszentrum für Gesundheit und Umwelt), KIT, LIP, PSNC, Ústav Informatiky Slovenská Akadémia Vied, Universitat Politècnica de València



DEEP HybridDataCloud receives funding from the European Union's Horizon 2020 research and innovation programme under agreement RIA 777435