ATTIVITÀ DI RICERCA E SVILUPPO SU TECNOLOGIE CLOUD

Marica Antonacci
Istituto Nazionale di Fisica Nucleare
Sezione di Bari

OVERVIEW

- Le attività di R&D ci consentono di studiare e testare soluzioni innovative per ampliare il portfolio dei servizi offerti alle comunità di utenti di ReCaS

- Principali progetti europei in cui la sede INFN di Bari è coinvolta

  - DEEP-Hybrid-DataCloud
  - eXtreme-DataCloud
  - EOSC-Hub
  - EOSC-Pillar

Questi 3 progetti costituiscono la prosecuzione e l'estensione di INDIGO-DataCloud che si è concluso il 30 Settembre 2017

FACILITARE L'ACCESSO ALLE RISORSE DI CALCOLO E STORAGE

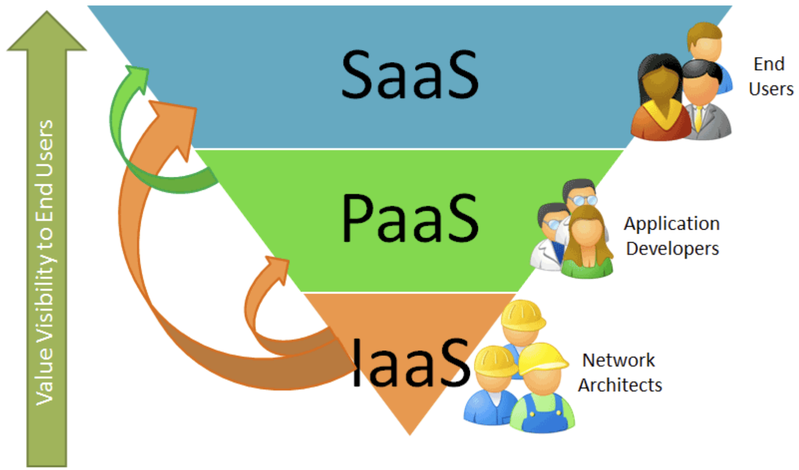
INDIGO-DATACLOUD

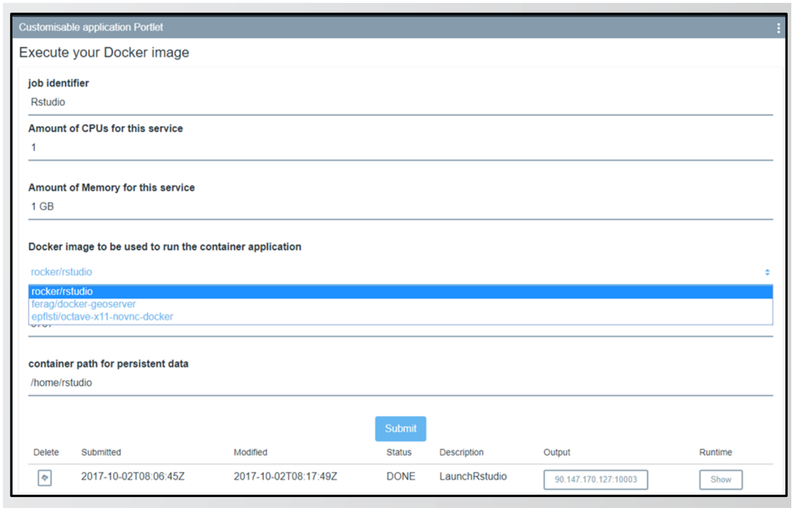
INDIGO - DataCloud
Better Software for Better Science

TOWARDS A SUSTAINABLE EUROPEAN PAAS-BASED CLOUD SOLUTION FOR E-SCIENCE

INDIGO - DataCloud develops an open source data and computing platform targeted at scientific communities, deployable on multiple hardware and provisioned over hybrid, private or public, e-infrastructures. By filling existing gaps in PaaS and SaaS levels, INDIGO-DataCloud will help developers, resources providers, e-infrastructures and scientific communities to overcome current challenges in the Cloud computing, storage and network areas.

https://www.indigo-datacloud.eu/





INFN

ORCHESTRAZIONE E AUTOMATIZZAZIONE

- **La PaaS di INDIGO offre funzionalità avanzate di orchestrazione** delle risorse in ambienti eterogenei:
  - Cloud private (Openstack, OpenNebula), pubbliche (Amazon, Azure, Google)
  - Piattaforme di orchestrazione di container (Apache Mesos)
- Consente il **deployment automatico** di infrastrutture virtuali (elastiche e/o ibride) o di cluster di servizi dockerizzati
  - **Automatizzazione** dell'installazione e della configurazione dei componenti sw
- **Accesso federato** tramite meccanismi di autenticazione/autorizzazione basati su OpenID-Connect (INDIGO IAM)

OASIS TOSCA

docker

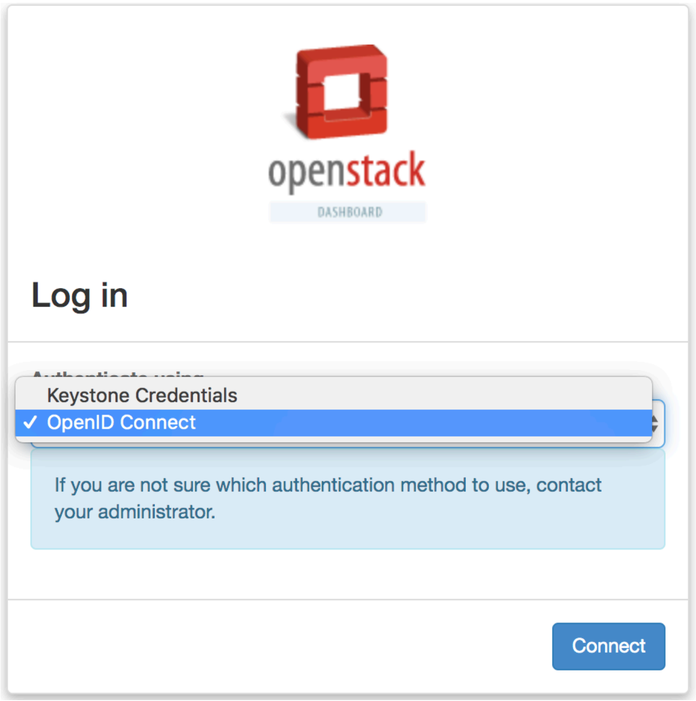
MESOS MARATHON

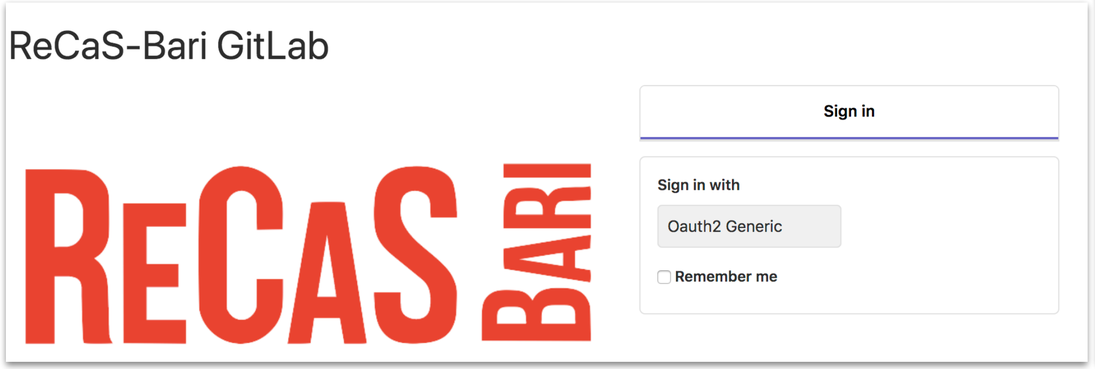
ANSIBLE

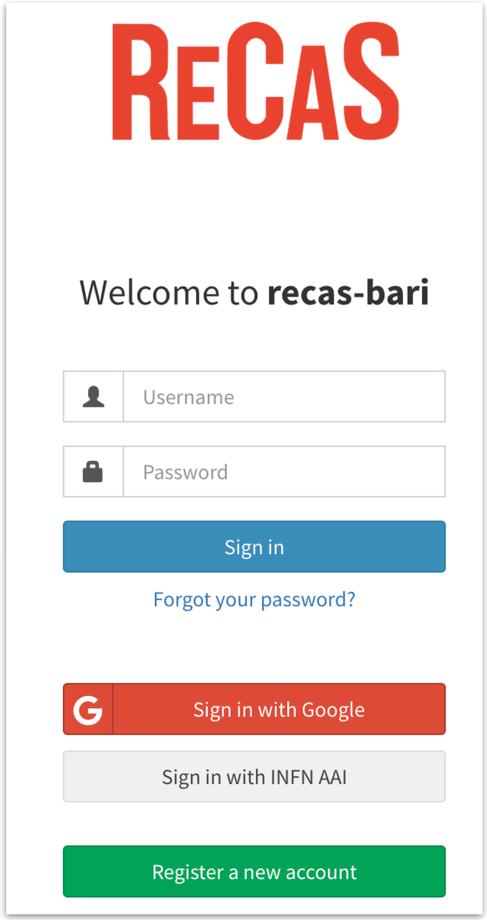
INFN

AUTENTICAZIONE FEDERATA

- **RECAS IAM è basato su INDIGO IAM**
- **Supporta differenti metodi di autenticazione: X509, OIDC, user/passwd**
- **Integrato con diversi servizi di ReCaS, p.e. Openstack e GitLab**

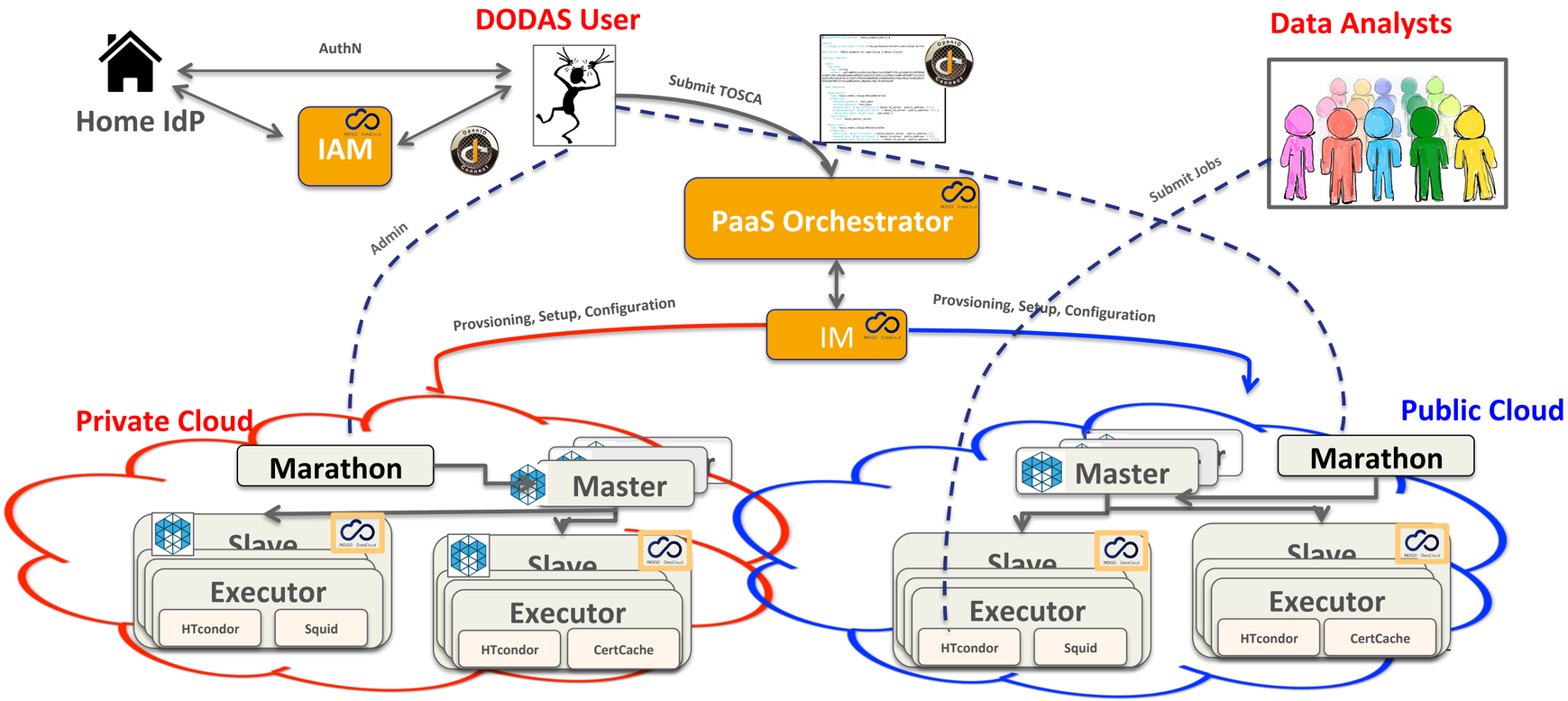






DODAS: How to effectively exploit heterogeneous clouds for scientific computations

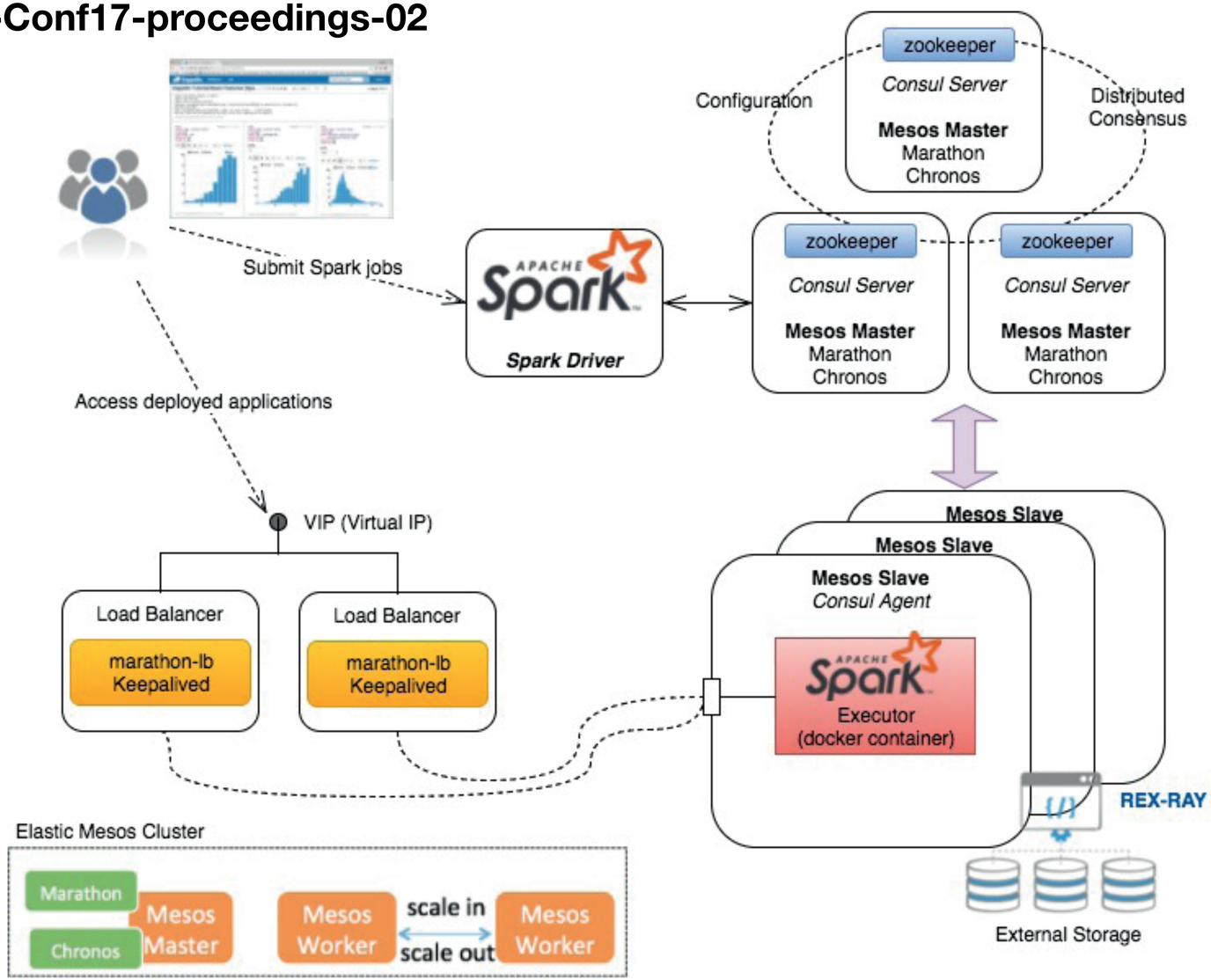
Presented at ISGC (International Symposium on Grids and Clouds) 2018
DOI: 10.22323/1.327.0024



On-demand instantiation of Condor-based clusters for HEP.

PIATTAFORME PER L'ANALISI DI BIG DATA ISTANZIATE ON-DEMAND TRAMITE LA PAAS DI INDIGO-DATACLOUD

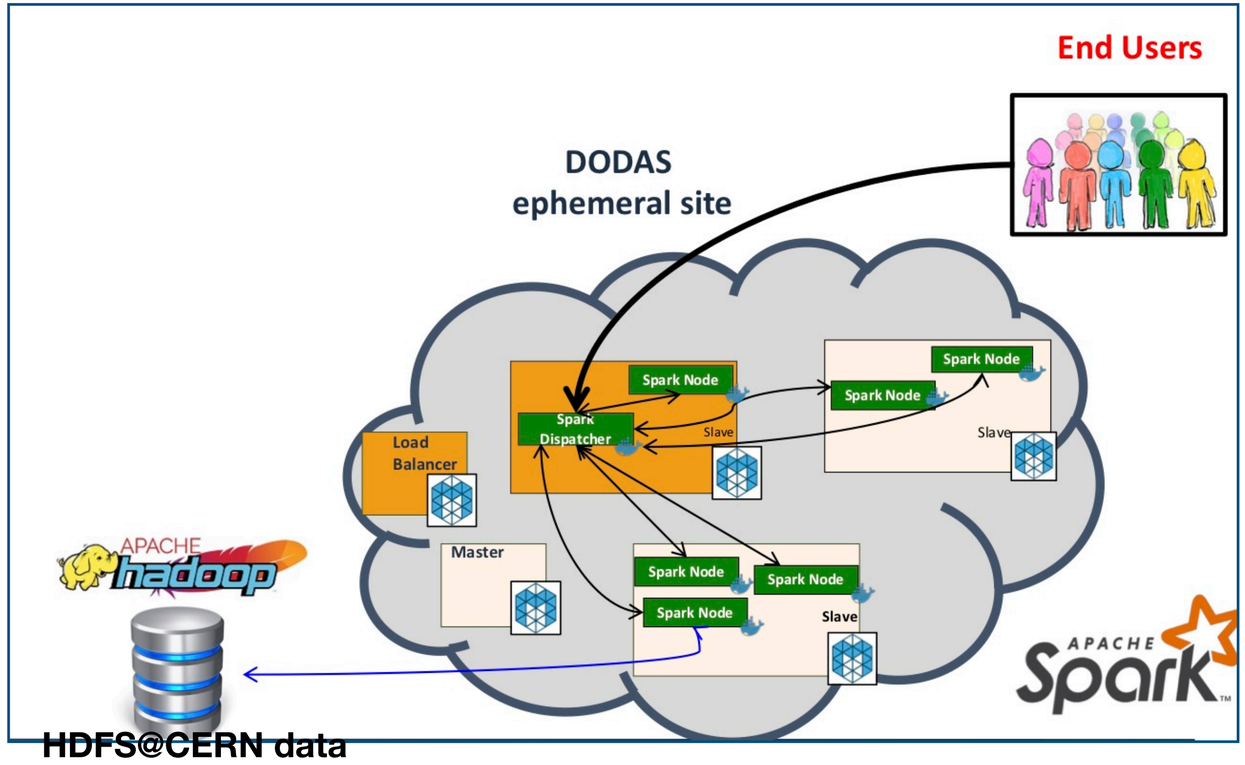
Conferenza GARR 2017
DOI: 10.26314/GARR-Conf17-proceedings-02



USING DODAS AS DEPLOYMENT MANAGER FOR SMART CACHING OF CMS DATA MANAGEMENT SYSTEM

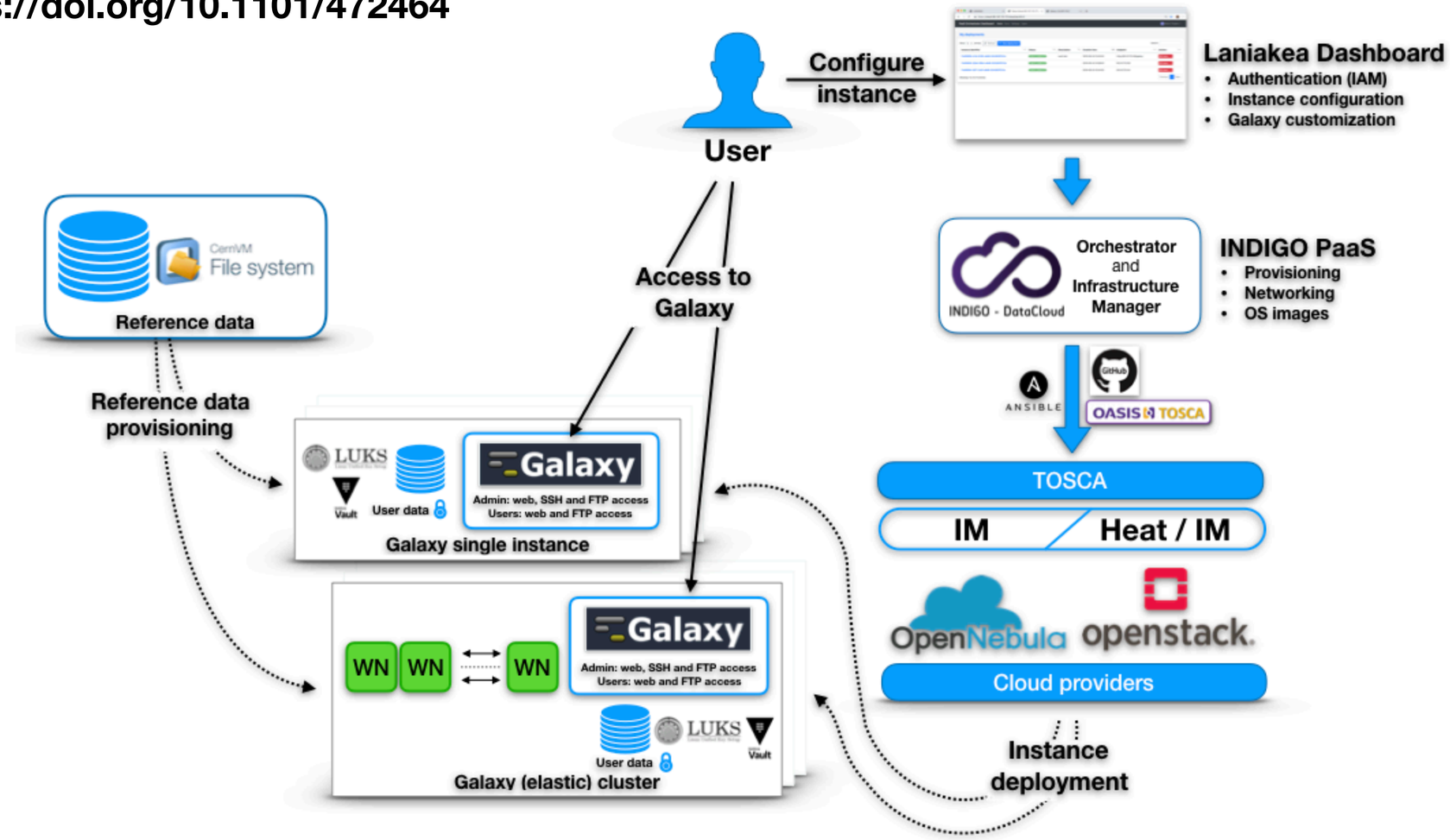
Presented at ACAT2019
(19th International Workshop on Advanced Computing and Analysis Techniques in Physics Research)

- AI solution for smart data caching
- The cache space should be dynamically populated with the most requested data
  - the data cache middleware is based on XRootD
- The inference system is based on the CMS-TensorFlow as a Service



LANIAKEA: GALAXY ON-DEMAND

Laniakea: an open solution to provide Galaxy "on-demand" instances over heterogeneous cloud infrastructures.
doi: https://doi.org/10.1101/472464



Laniakea is being developed by ELIXIR-IIB, the italian node of ELIXIR

The production phase of the ELIXIR-IT Laniakea@ReCaS service will start in the second half of 2019

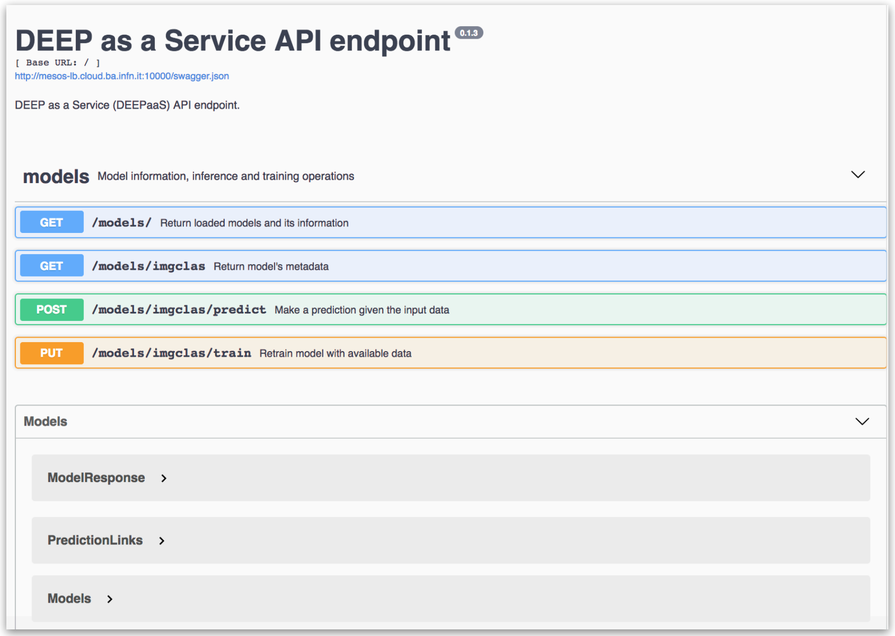
DEEP LEARNING AS A SERVICE

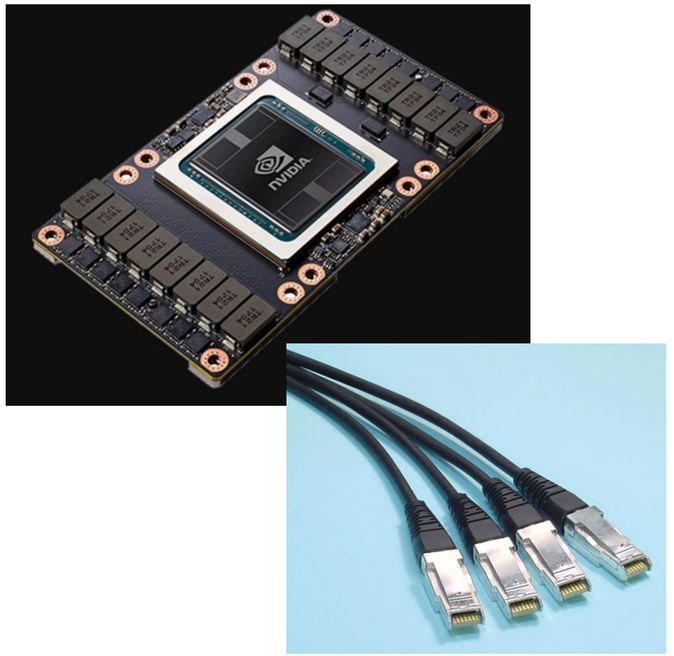
DEEP-Hybrid-DataCloud

DEEP Hybrid DataCloud project aims to **support intensive computing techniques** that require specialized HPC hardware, like GPUs or low-latency interconnects, to explore very large datasets.

Under the common label of **DEEP as Service**, the project will provide a set of building blocks that enable the easy development of applications requiring cutting-edge tecniques: **artificial intelligence (machine learnig and deep learning), parallel post-processing of very large data, and analysis of massive online data streams.** These services will be deployed in the project testbed, offered to the research communities linked to the project through pilot applications, and **integrated under the EOSC framework.**

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





SERVIZI PER ML

- **DEEP-Hybrid-DataCloud mette a disposizione servizi utili per tutte le fasi del ciclo di Machine Learning:**
  - Creare un modello da zero o partendo da uno esistente
  - Addestrare, testare e validare il modello
  - Gestire il deployment del modello e la sua pubblicazione come “servizio on demand”

- **Tecnologie:**
  - Docker container
  - Accesso trasparente alle GPU
  - Deployment ibridi
  - Integrazione con HPC
  - Architettura serverless
  - Marketplace con modelli ready to use
  - Standard APIs per le fasi di training/testing/inference

SOLUZIONI PER IL DATA LAKE

eXtreme DataCloud

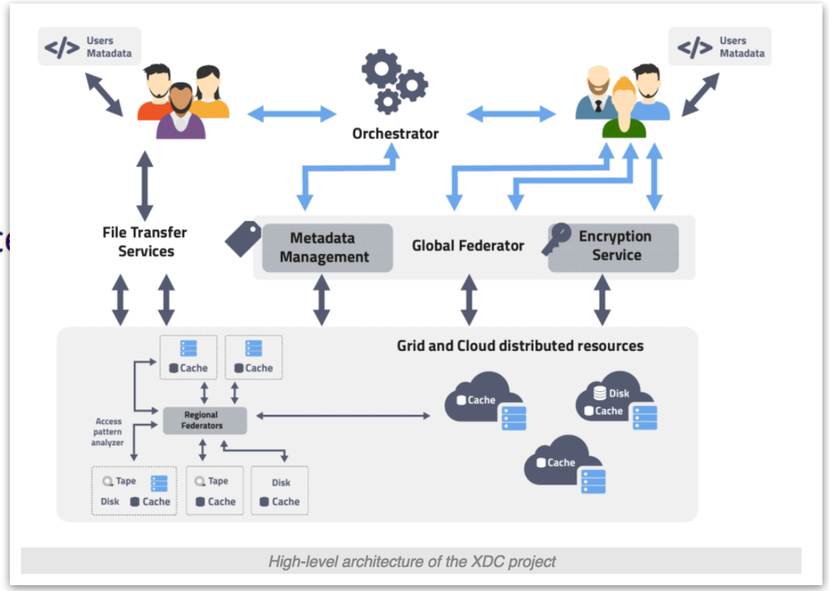
http://www.extreme-datacloud.eu

Home About Materials Software Use cases News and events Contact

The eXtreme DataCloud (XDC) is a EU H2020 funded project aimed at developing scalable technologies for federating storage resources and managing data in highly distributed computing environments. The services provided will be capable of operating at the unprecedented scale required by the most demanding, data intensive, research experiments in Europe and Worldwide.

The main high-level topics addressed by the project include:

- federation of storage resources with standard protocols
- smart caching solutions among remote locations
- policy driven data management based on Quality of Servic
- data lifecycle management
- metadata handling and manipulation
- data preprocessing and encryption during ingestion
- optimized data management based on access patterns.



High-level architecture of the XDC project

INFN

Verso la European Open Science Cloud

- EOSC (European Open Science Cloud), l'infrastruttura per la condivisione di dati e servizi al servizio della scienza
- Alcune soluzioni e servizi realizzati nell'ambito dei progetti INDIGO, DEEP, XDC, etc. sono parte del marketplace di EOSC, p.e.:
  - DODAS Thematic Service
  - DEEPaaS training facility
- Galaxy on-demand è uno degli use-case di EOSC-Pillar

GRAZIE PER L'ATTENZIONE