CloudVeneto and INFN Cloud: an overview

Marco Verlato (INFN Padova)

1st AI-INFN Advanced Hackathon

26–28 November 2024

University of Padua, Complesso Paolotti

A little of history

From the Grid age...

- EU-DataGrid (2001-2004)
- EU-DataTAG (2002-2004)
- EGEE-I,II,III (2004-2010)
- OMII-Europe (2006-2008)
- EMI (2010-2013)

WLCG Worldwide LHC Computing Grid

... to the Cloud age

- EGI-INSPIRE (2010-2014)
- Open City Platform (2014-2016)
- EGI-ENGAGE (2015-2017)
- INDIGO-DataCloud (2015-2017)
- eXtreme DataCloud (2017-2020)
- EOSC-HUB (2018-2021)
- EOSC-FUTURE (2021-2024)

2001

TODAY

Technology evolution

- Xen (2003)
- KVM (2006)
- Amazon EC2 (2008)
- OpenStack (2010)
- Docker (2013)
- Kubernetes (2015)

Padova and LNL

- INFN-Padova and Legnaro National Labs (LNL) are ~10 km far away
- Longstanding collaboration as WLCG Tier-2 for ALICE and CMS experiments
- A single Tier-2 with resources distributed in the two data centers
- 80 Gbps dedicated network connection (soon 2x100 Gbps)
- Sharing of infrastructure, hardware and human resources

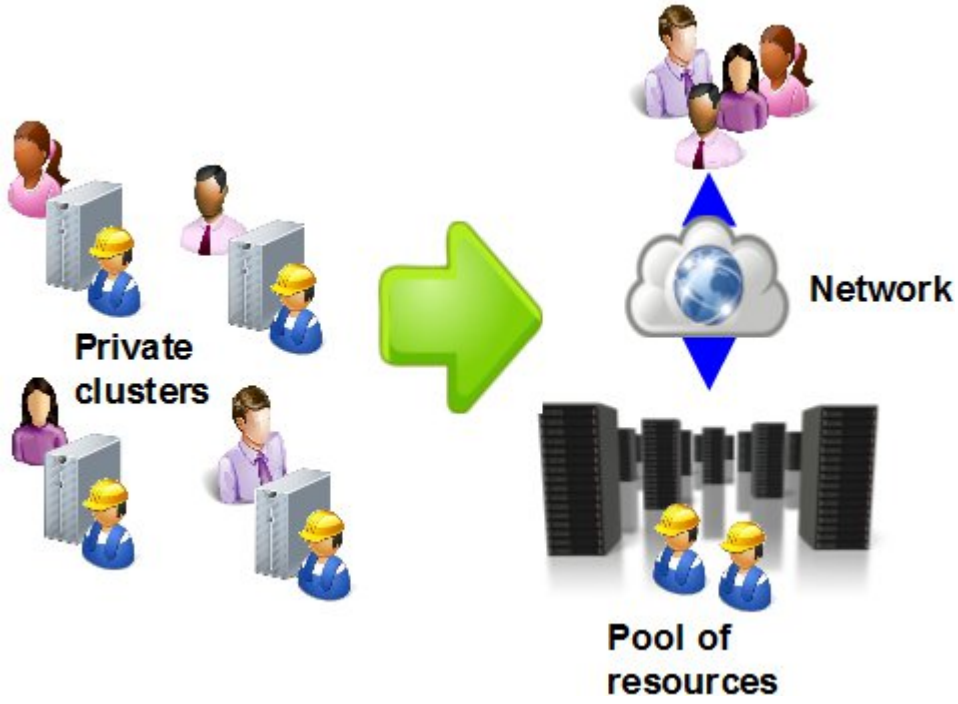
Grid was a success, but...

- Well known entry barriers affecting several smaller sized physics experimental teams:
  - ✓ Getting and managing X509 personal certificates
  - ✓ Limited flexibility of computing environments
  - ✓ Lack of interactivity
- These groups tend to buy independently their own clusters to satisfy their computing needs, leading to the following drawbacks:
  - ✓ A lot of heterogeneous small sized clusters in the same data center
  - ✓ These cluster are often underutilized, while close to the deadlines are insufficient
- Low overall efficiency and high system administration cost

Cloud is the solution

- Higher efficiency and elasticity at lower costs:
  - ✓ A single large IaaS cloud computing facility centrally managed
  - ✓ On-demand self-service elastic provisioning of resources
  - ✓ No X509 certificates issue, Federated Identity management systems are enabled
  - ✓ Users can choose Operating System, #CPUs, RAM and Storage sizes better suited for their computing tasks
  - ✓ Interactive access to Virtual Machines is allowed

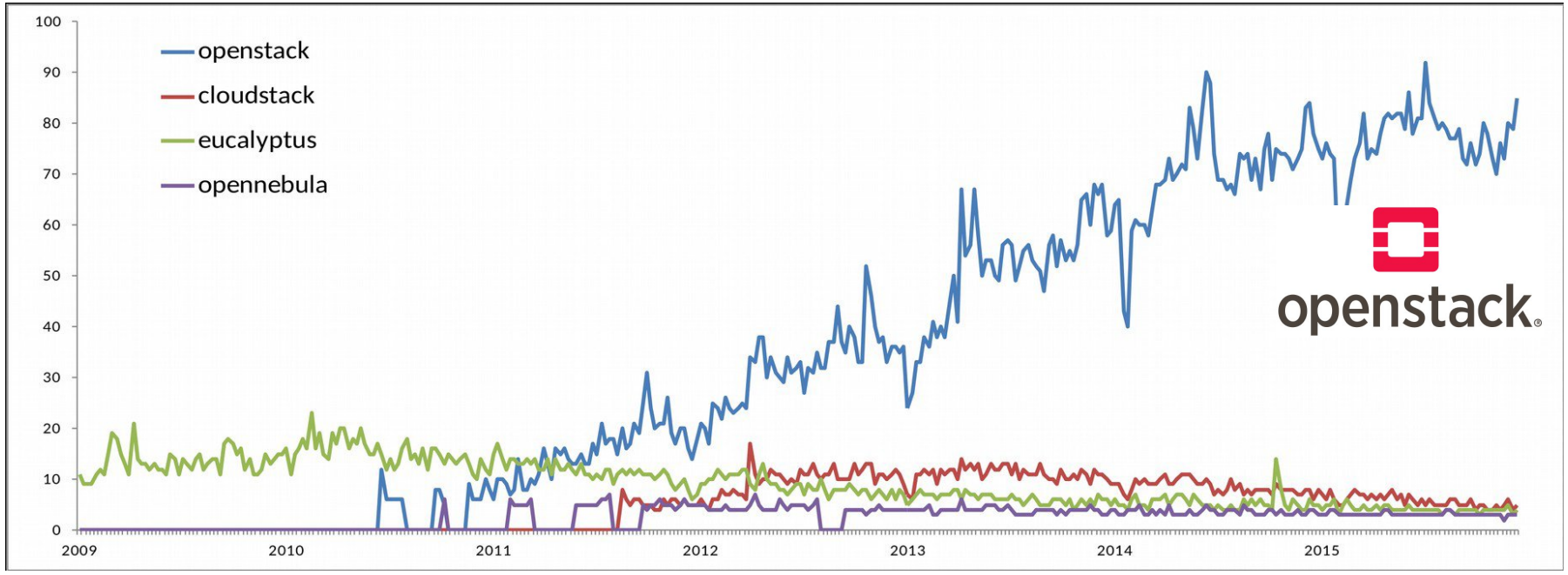
- The experimental groups buy a quota of this large shared computing facility instead of buying their own physical cluster



Open source solutions

In 2013 we chose OpenStack as cloud middleware

- Open source software, open development process, open community with strong support by ICT industry
- Worldwide continuously increasing dimension and product in fast and deep evolution
- Modular architecture
- Massive adoption in our community (e.g. at CERN)



Merging two clouds

- Cloud Area Padovana infrastructure in production at the end of 2014
- Resources physically distributed among Padova and LNL data centers, and initially funded by INFN only
- Additional funds from UniPD + 10 scientific departments led to another OpenStack based cloud infrastructure in 2015, hosted in the same Padova data center
- The two infrastructures merged in 2018 to form CloudVeneto

Deployment

CloudVeneto deploys basic OpenStack components for Compute, Storage, Networking, Identity, Images, Orchestration services, available to users through a web dashboard (Horizon)

Compute

NOVA	Compute Service
ZUN	Containers Service

Storage

SWIFT	Object store
CINDER	Block Storage
MANILA	Shared filesystems

Networking

NEUTRON	Networking
OCTAVIA	Load balancer
DESIGNATE	DNS service

Shared Services

KEYSTONE	Identity service
PLACEMENT	Placement service
GLANCE	Image service
BARBICAN	Key management

Hardware Lifecycle

IRONIC	Bare Metal Provisioning Service
CYBORG	Lifecycle management of accelerators

Application Lifecycle

MASAKARI	Instances High Availability Service

Orchestration

HEAT	Orchestration
MISTRAL	Workflow service
ZAQAR	Messaging Service
BLAZAR	Resource reservation service
AODH	Alarming Service

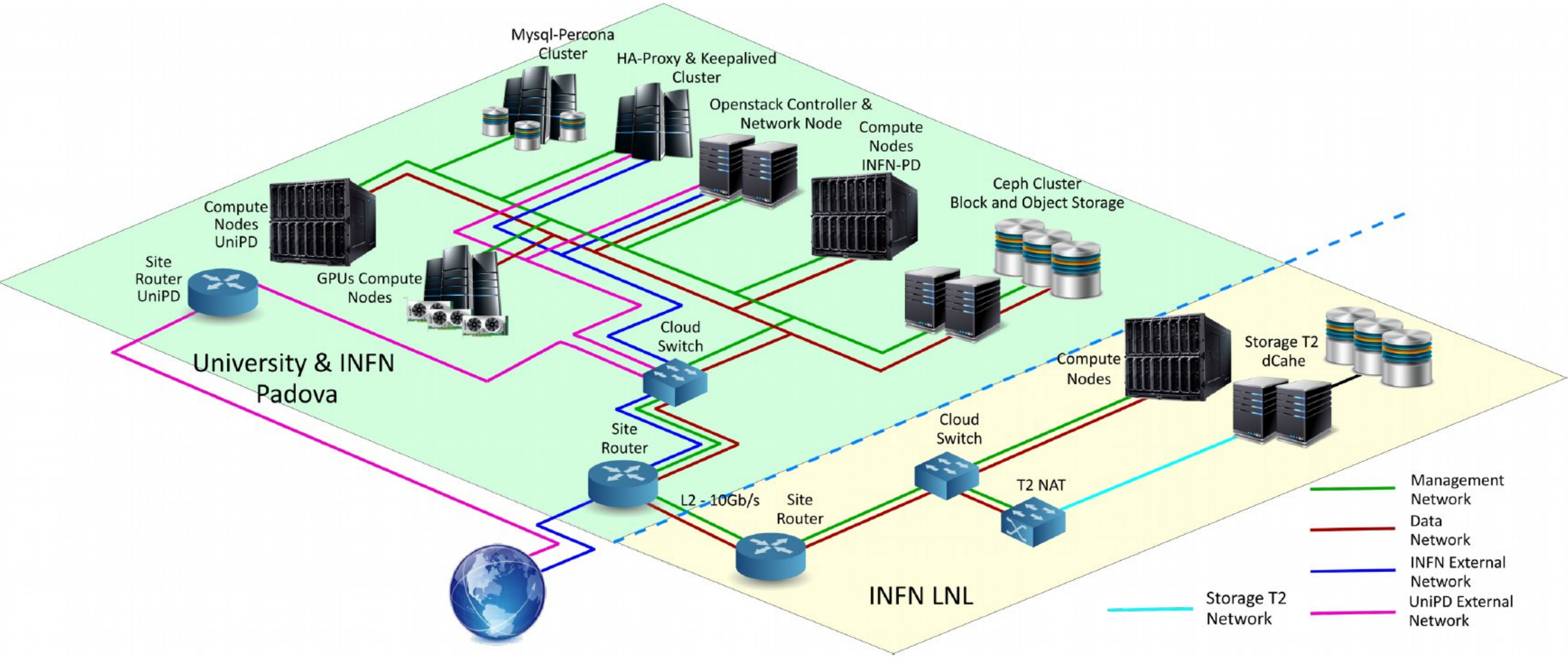
Web frontends

HORIZON	Dashboard
SKYLINE	Next generation dashboard (emerging technology)

Workload Provisioning

MAGNUM	Container Orchestration Engine Provisioning
TROVE	Database as a Service

Resource distribution layout



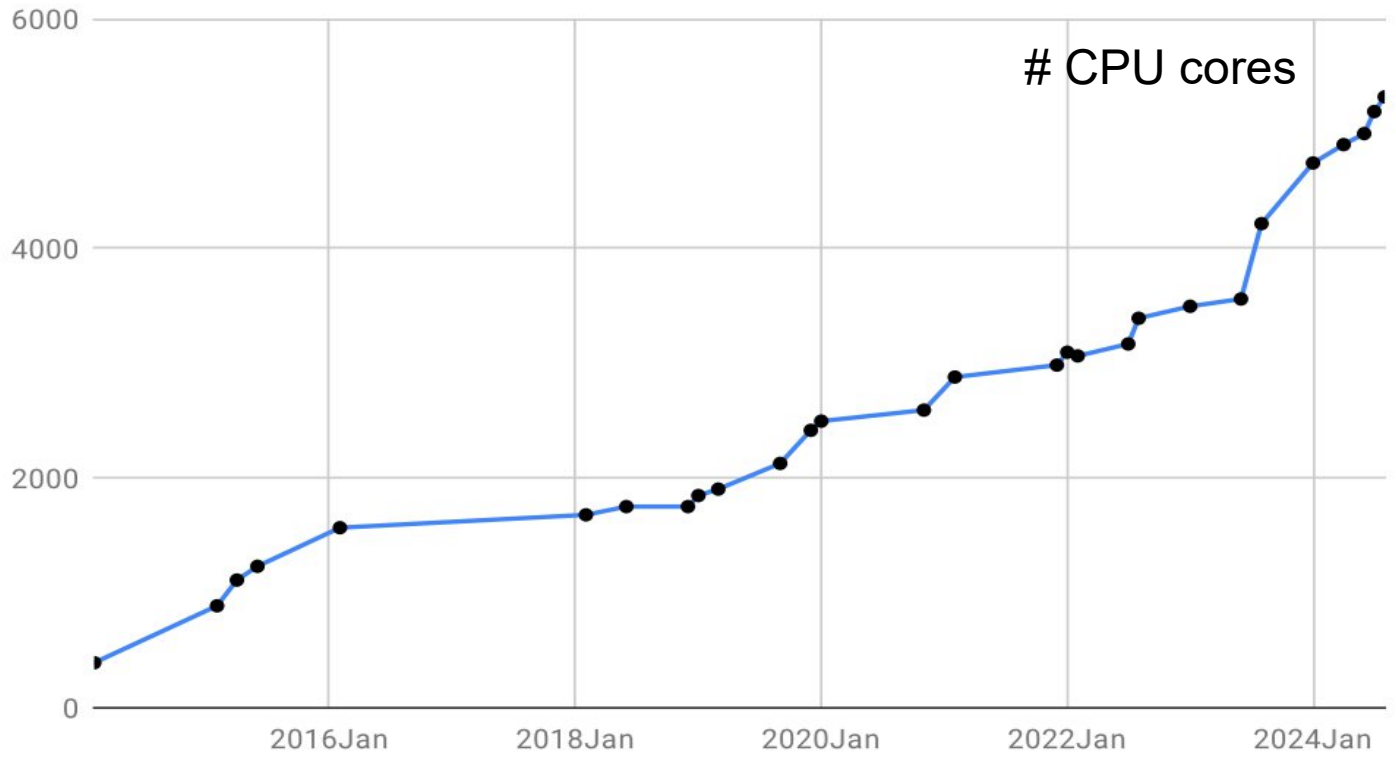
Network configuration

- By default the VM instances are on a private network
  - ✓ Outbound connectivity
  - ✓ VM accessed via ssh from INFN-PD or LNL LANs
  - ✓ ... or through a gate
- If needed, VMs can access to Tier-2 storage network
- Public IPs given only for specific cases
  - ✓ INFN public IPs
  - ✓ UniPD public IPs
- Public IPs scanned regularly to check for vulnerabilities through Greenbone

Resource capacity

- Hardware resources have undergone continuous expansion and renewal over the years
- Current total capacity of storage and computing resources:

  - ✓ Compute nodes: **68**
  - ✓ CPU cores (in HT): **5320**
  - ✓ RAM: **28 TB**
  - ✓ GPUs: **29**
  - ✓ Storage: **2.0 PiB** (raw) (via ceph cluster)



User view

- Infrastructure as a Service (IaaS) model
  - ✓ The user can instantiate, manage and use its own computing infrastructure built with Virtual Machines (VM), storage and network
    - pre-defined or custom OS images
    - many available flavors (to choose #VCPUs, RAM, ephemeral storage size)
    - data on storage volumes attached to VMs (block storage) or object storage (ceph-rgw exposing both S3 and Swift interfaces)
- Higher level services are also available
  - ✓ Resource orchestration
  - ✓ Elastic batch clusters support (based on HTCondor and Slurm)
  - ✓ Container as a Service (CaaS) platform (based on Kubernetes)

Projects

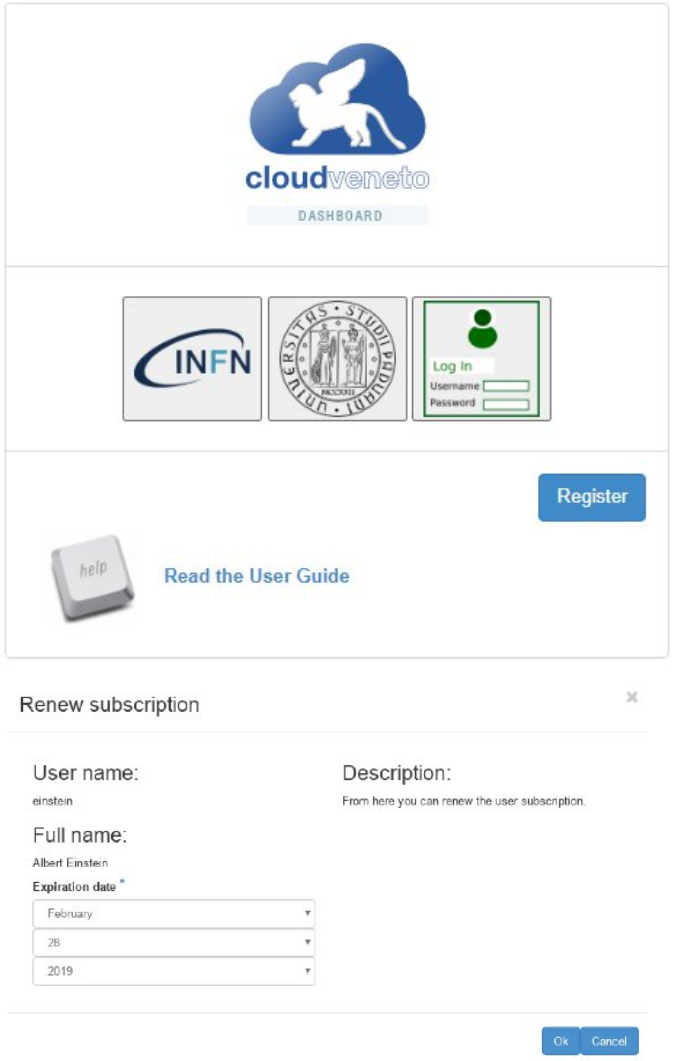
CloudVeneto resources and users are grouped into projects:

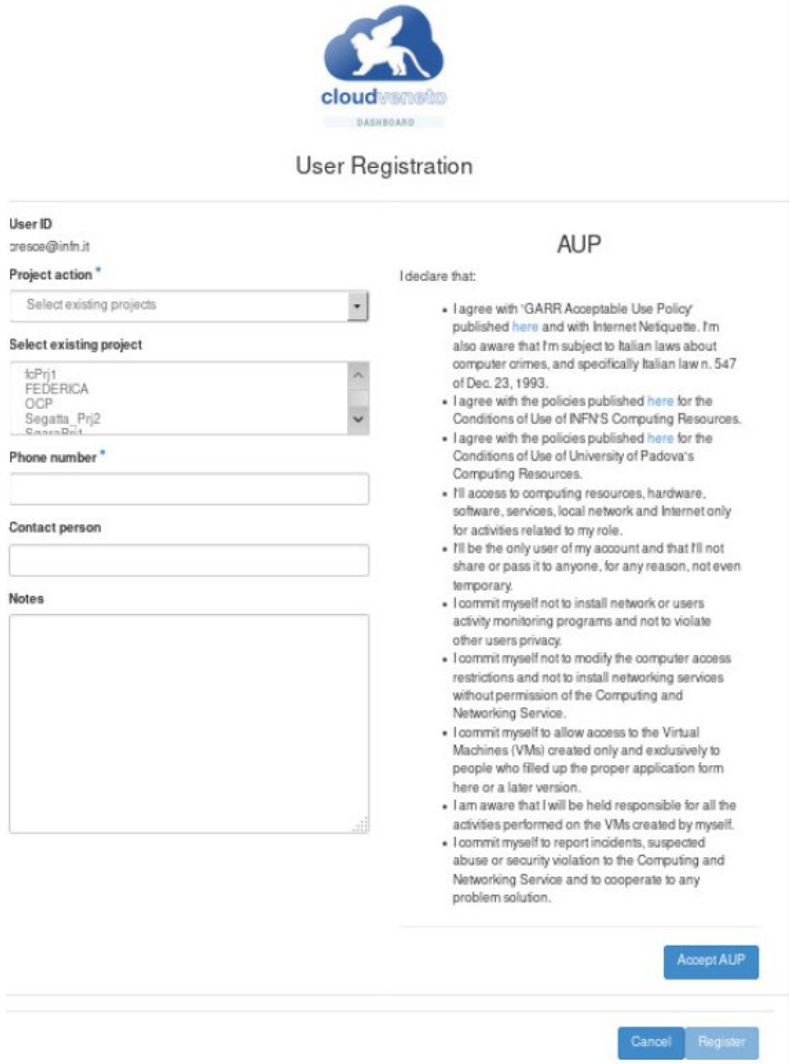
- Typically an experiment/working group/research project
  - ✓ project creation must be approved by the Department contact person
  - ✓ each project has a (renewable) expiration date
  - ✓ each UniPD/INFN labelled project has a quota on UniPD/INFN paid resources
- A user can join one or more projects
  - ✓ each project has a manager who approve/refuse users affiliation requests
  - ✓ user affiliation to a given project has a (renewable) expiration date
  - ✓ users who need to use INFN IT resources must met some prerequisites to be compliant with INFN rules

Original developments

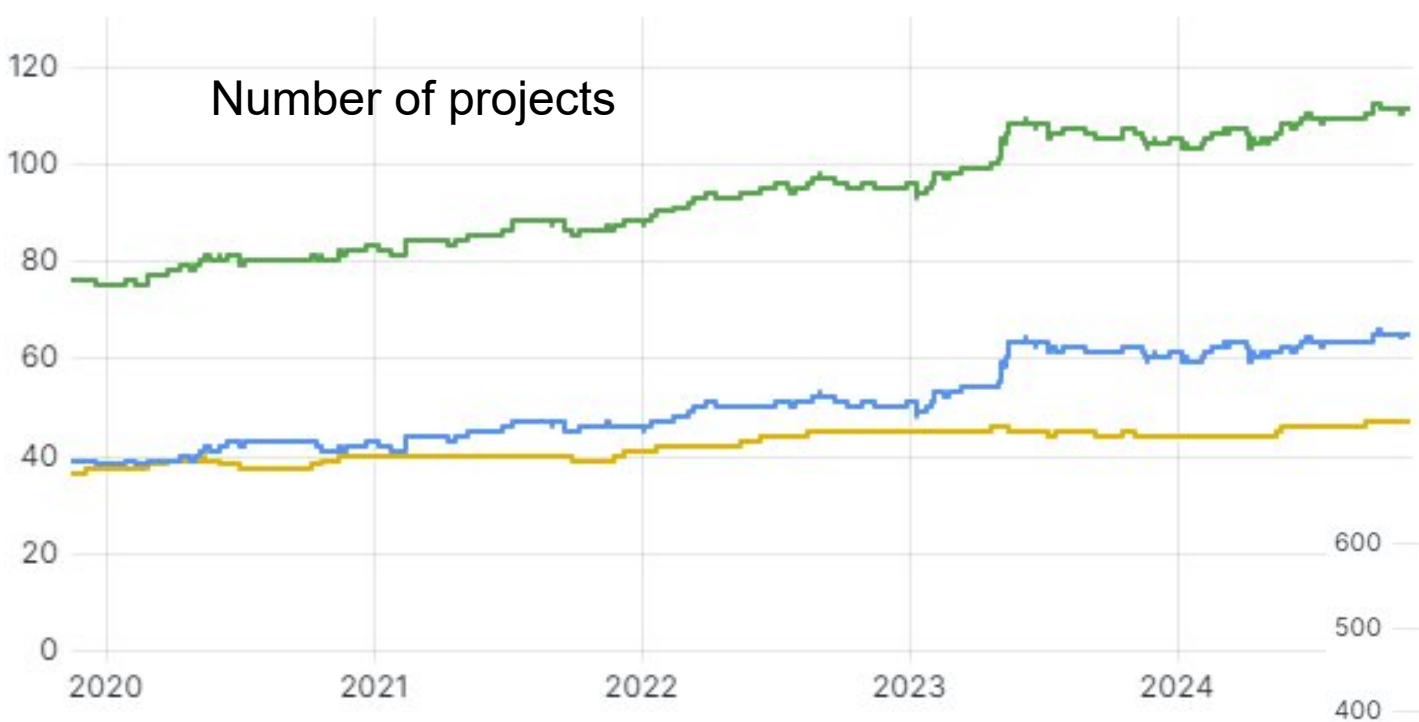
Some work needed to:

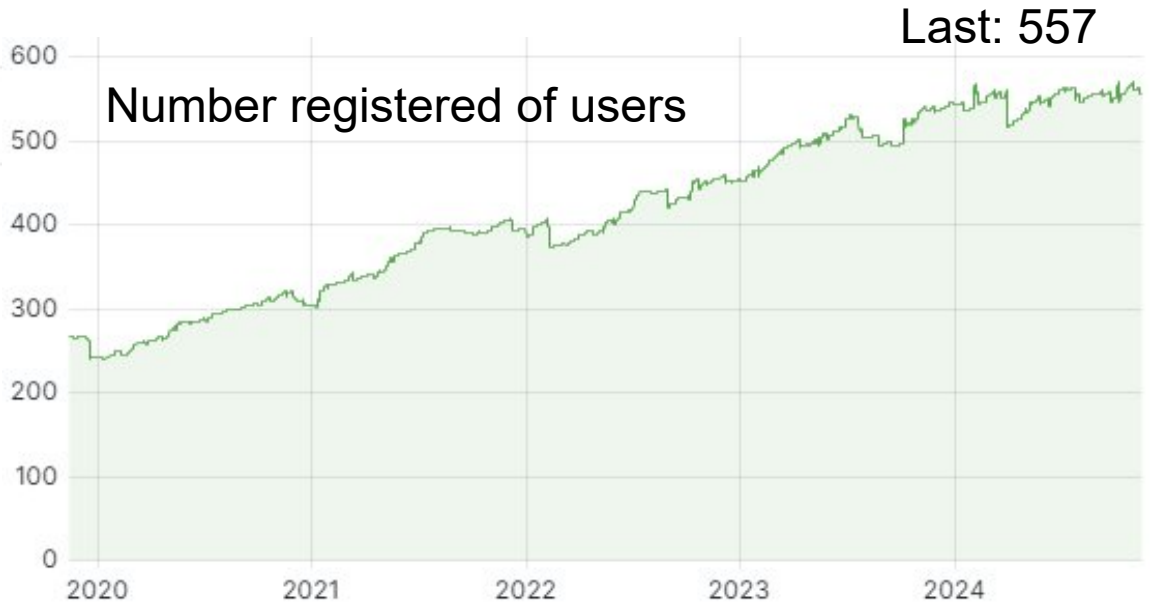
- Support user authentication through Identity Providers
- Support the workflows to:
  - ✓ register
  - ✓ approve requests
  - ✓ renew subscriptions
  - ✓ create new users
  - ✓ create new projects
- Everything automated and integrated in the Horizon Dashboard



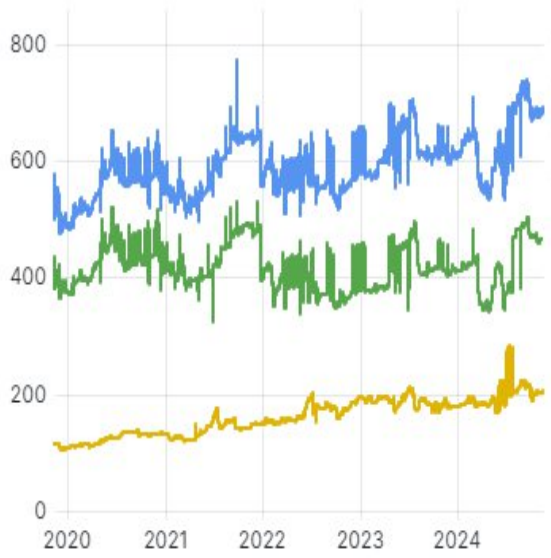


Project & user growth (last 5 years)





Usage growth (last 5 years)



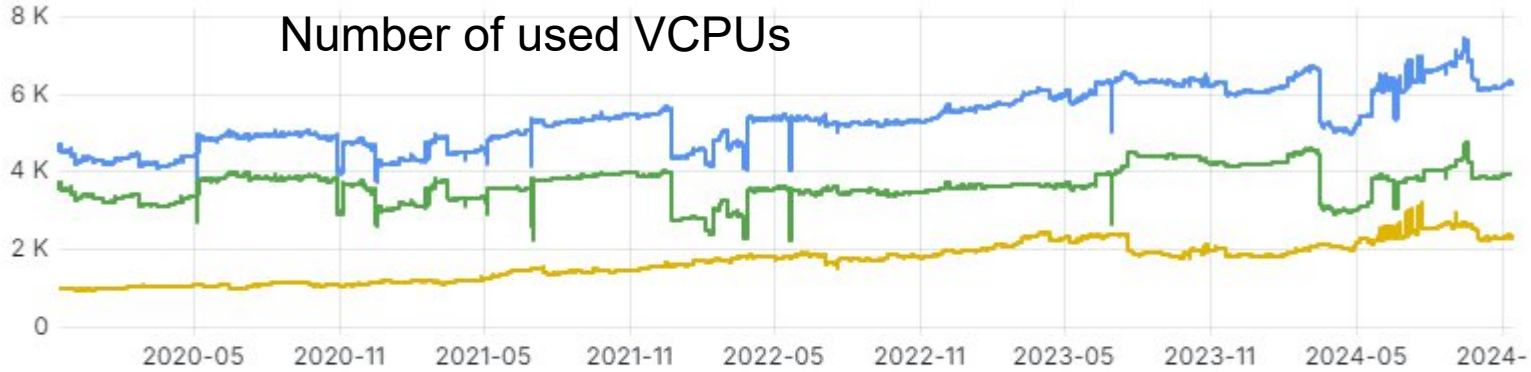
Name	Last *
cloudvenetoprod_org_usage.instances { org: INFN }	467
cloudvenetoprod_org_usage.instances { org: Unipd }	207
cloudvenetoprod_total_usage.instances	686

Number of VMs

Name	Last *
cloudvenetoprod_org_usage.ram { org: INFN }	10.0 TiB
cloudvenetoprod_org_usage.ram { org: Unipd }	6.14 TiB
cloudvenetoprod_total_usage.ram	16.2 TiB

RAM

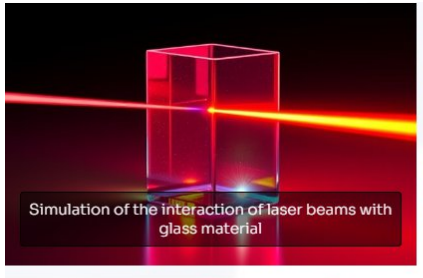
Number of used VCPUs

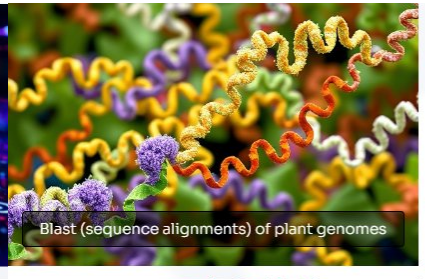


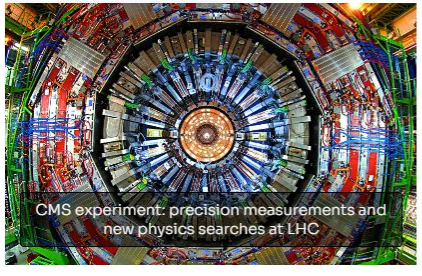
Name	Last *
cloudvenetoprod_org_usage.cores { org: INFN }	3.92 K
cloudvenetoprod_org_usage.cores { org: Unipd }	2.39 K
cloudvenetoprod_total_usage.cores	6.33 K

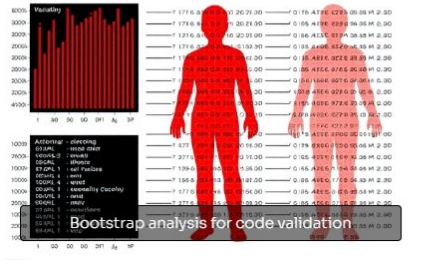
Use cases

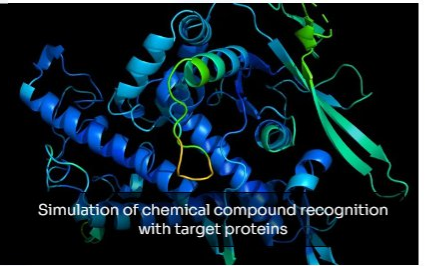
- Several scientific computing applications in diverse disciplines
  - ✓ see experiences collected at www.cloudveneto.it
- ... but also other use cases such as teaching
  - ✓ Master degree in “Physics of Data” at Department of Physics & Astronomy
  - ✓ “Big Data” course at Department of Information Engineering

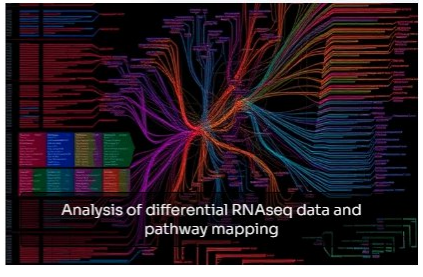












Governance (new)

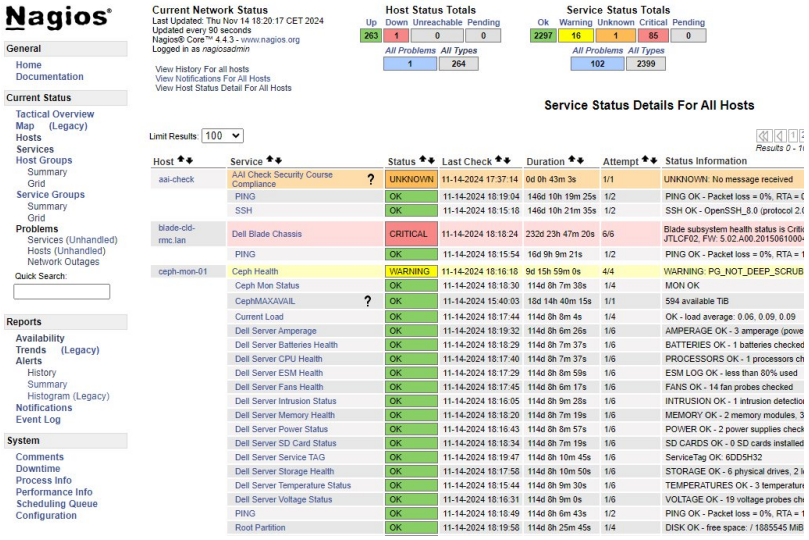
- Scientific Board (~yearly)
  - ✓ Representatives from each UniPD Dep. + INFN-PD and INFN-LNL appointed by the Directors
  - ✓ Defines the guidelines for the development of the activities carried out on the infrastructure
  - ✓ Coordinated by the Management Board coordinator
- Management Board (~monthly)
  - ✓ 5 members, electing a coordinator, + the Technical Board coordinator
  - ✓ Supervises all non-technical project activities, e.g. resource allocation, access policies, special requests from new experiments/groups etc.
- Technical Board (~weekly)
  - ✓ A dozen of technologists/technicians from UniPD and INFN in charge of managing the operations of the infrastructure
  - ✓ Coordinator appointed by the Management board
  - ✓ Plans and implement the hardware and software maintenance and evolution of the infrastructure, including the users resource allocation according to the policies defined by SB/MB
  - ✓ Does not install and configure users scientific software and their VM instances

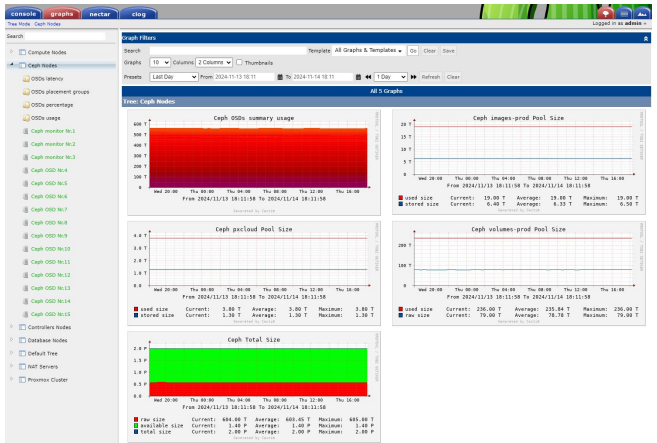
Operations

- Technical team of 14 people dedicating a (often very small) fraction of their time to maintain and support CloudVeneto
- Weekly user and infrastructure support rotation
  - ✓ Pre-check, oversee user registration and related processes
  - ✓ Help Desk —> user support mailing list support@cloudveneto.it
  - ✓ Malfunctioning investigation —> monitoring tools
  - ✓ Sysadmin support to VM mainly left to local computing services
  - ✓ Documentation at www.cloudveneto.it —> user guide
- Weekly meetings to coordinate the technical activities
  - ✓ To fix hardware/software/network or user issues
  - ✓ To plan and implement hardware/software/middleware updates

Monitoring

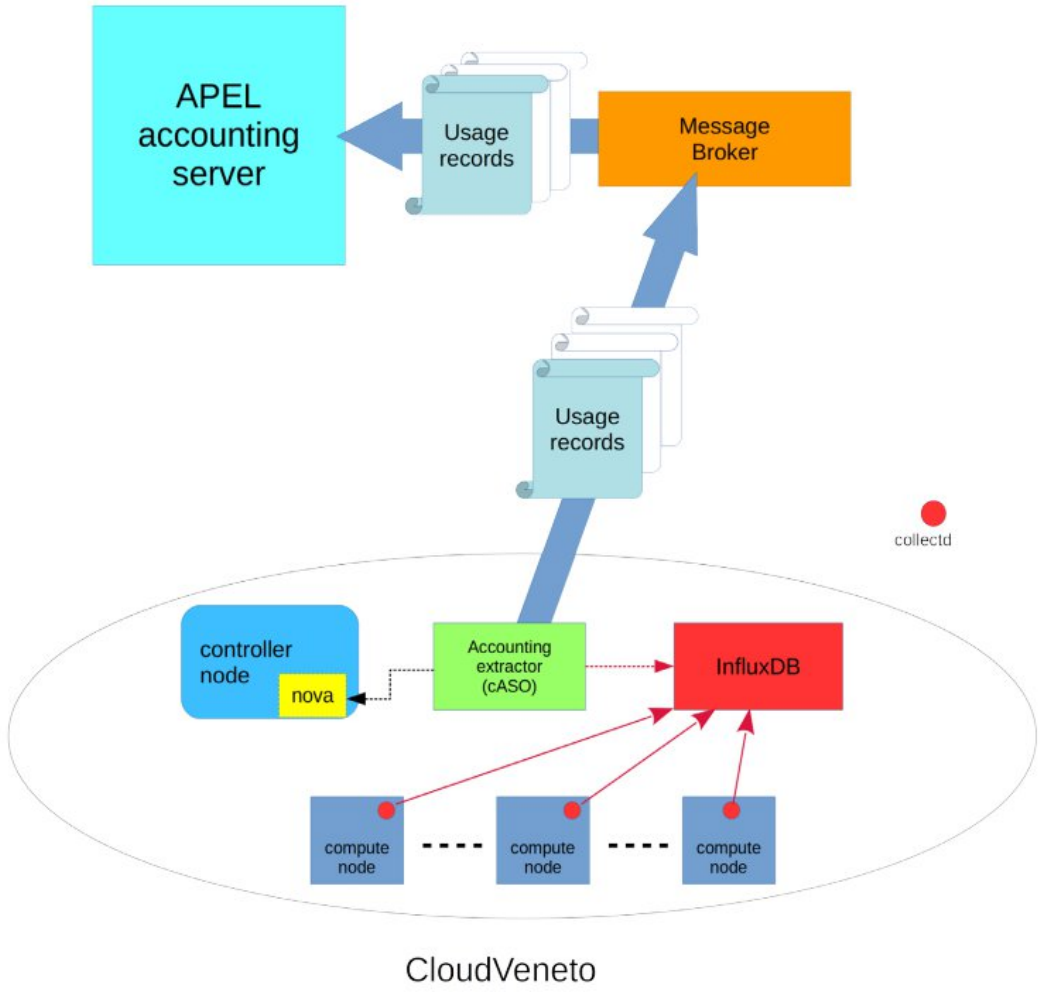
Leveraging some technologies adopted by Tier-2 too





Accounting

- True CPUTime not available from OpenStack API
- Accounting tool from EGI modified to satisfy CloudVeneto requirements
- Each compute node instrumented via collectd (virt plugin)
- Collectd collects CPUTime consumed by each instance, and send these data to a InfluxDB instance
- cASO modified to get CPUTime information querying the InfluxDB
- Not needed to install anything on the cloud virtual machines
- **This solution has been later adopted by INFN Cloud**



Changes wrt EGI FedCloud architecture are in red

Auditing

Security auditing is challenging in cloud environment

- Even more complex for our peculiar network set up
- Typical security incident: something bad originated from IP a.b.c.d at time YY:MM:DD:hh:mm
- A procedure was defined to manage security incidents:
  - ✓ Given the IP a.b.c.d, to find the VM private IP
  - ✓ Given the VM private IP, to find the MAC address
  - ✓ Given the VM MAC address, to find the UUID
  - ✓ Given the VM UUID, to find the owner
- The above workflow is possible by using specific tools (netfilter.org ulogd, CNRS os-ip-trace) and archiving all the relevant log files
- It allows to trace any internet connection initiated by a VM on the cloud, even if in the meantime it was destroyed

OpenStack updates

- Currently using OpenStack Yoga version, ready to update to Caracal
- So far “fast forward update” mode:
  - ✓ An update every ~ 1.5 years
  - ✓ Updating 4-5 versions of OpenStack at once
  - ✓ Everything tested in a dedicated devel cloud infrastructure
- In the future, update every 2 versions (now supported by OpenStack)
- Automatic provisioning and configuration through Foreman and Puppet customized modules

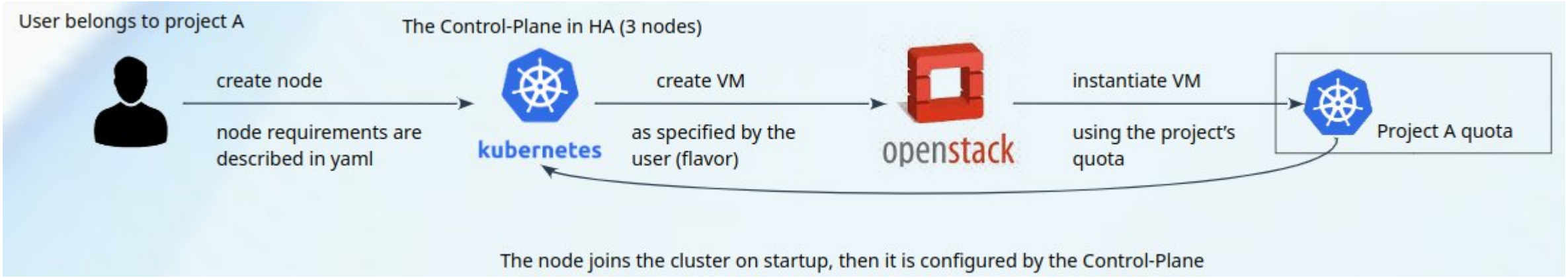
Foreman

puppet

Innovative solutions

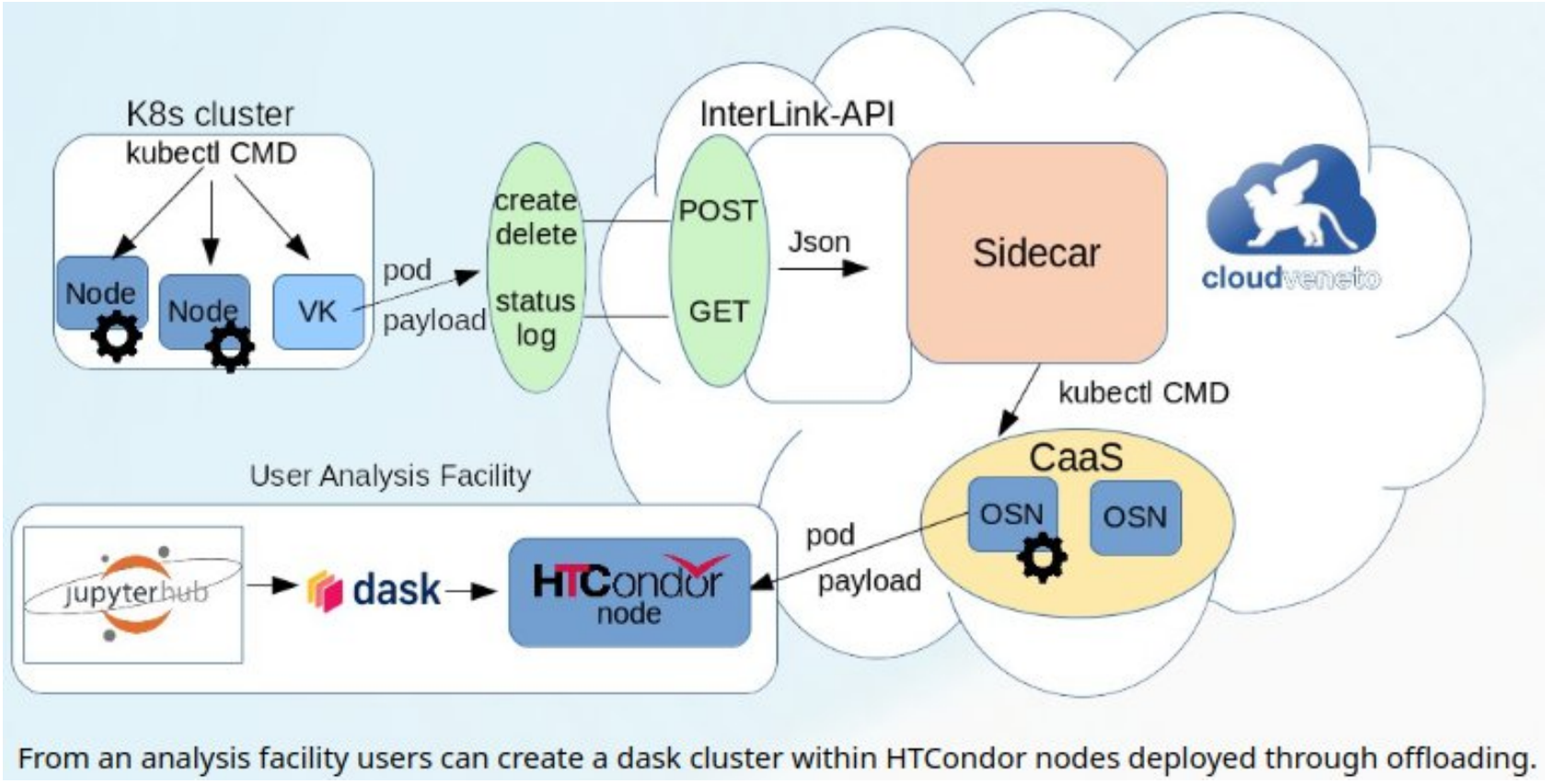
Container-as-a-Service (CaaS) offers a centralized Kubernetes Control Plane:

- users create VMs within their CloudVeneto project that become nodes of a K8s cluster where they can deploy their container based applications
- users can keep the nodes private or share them within the project
- users don’t have to deal with the complexity of administering a K8s cluster
- isolation at both user and node levels achieved by enhancing the K8s security layer by implementing multi-tenancy management through webhooks



CaaS for AI_INFN

- Offloading allows the execution of containerized code across distributed and heterogeneous computational environments
- interLink impl. for the interTwin project:
  - ✓ Virtual Kubelet
  - ✓ InterLink API
  - ✓ interLink sidecar
- We developed a custom CaaS interLink sidecar
- Solution validated for CMS High Rate analysis Facility and AI_INFN



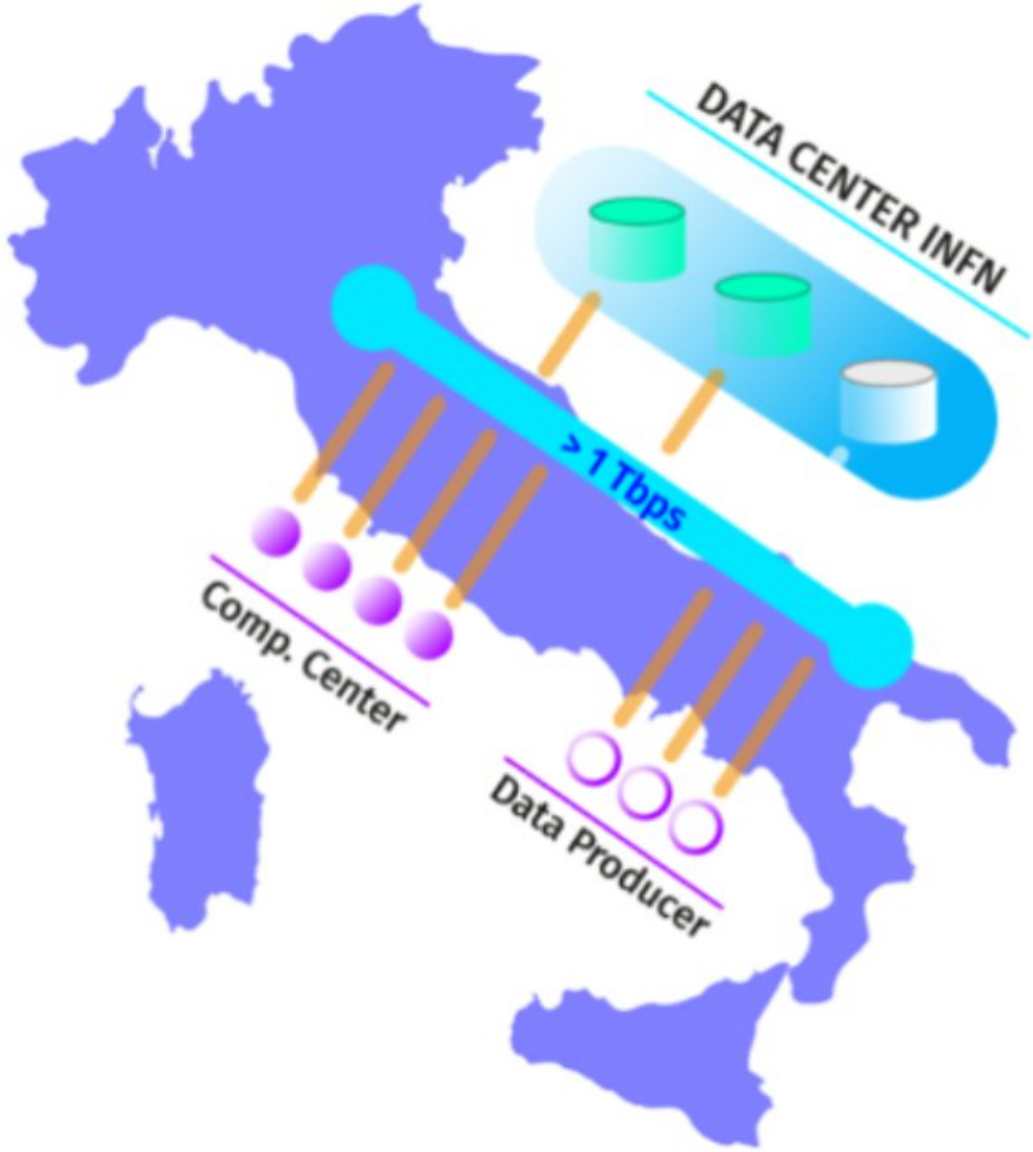
From an analysis facility users can create a dask cluster within HTCondor nodes deployed through offloading.

From local experiences to INFN Cloud

- Not only CloudVeneto, but several INFN sites have been investing for many years in cloud computing infrastructures (CNAF, Bari, Torino, ...)
- To optimize the use of available resources and expertise, in 2020 INFN decided to implement a national cloud infrastructure for research
  - ✓ as a federation of existing distributed infrastructures extending them if necessary in a transparent way to private and commercial providers
  - ✓ as an “user-centric” infrastructure making available to the final users a dynamic set of services tailored on specific use cases
  - ✓ leveraging the outcomes of several national and European cloud projects where INFN actively participated
- INFN Cloud officially available to users in March 2021

Federation

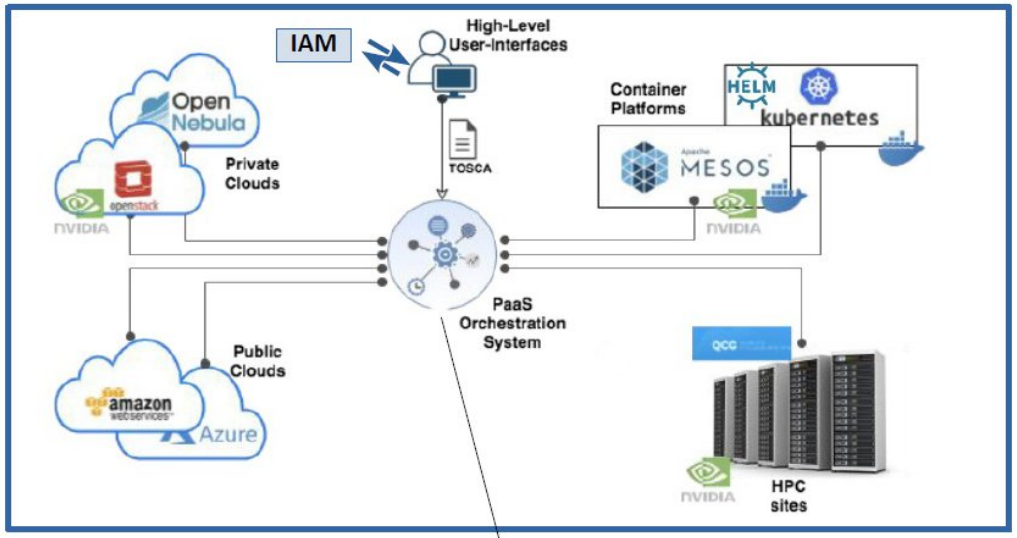
- INFN Cloud is a federation of pre-existing infrastructures:
  - ✓ a Backbone of made up of two closely linked federated sites: Bari and CNAF
  - ✓ a scalable set of satellite sites
    - expand the resources offered by the backbone
    - geographically distributed across Italy (Recas-Bari, Catania, CNAF, CloudVeneto, Napoli...)
    - heterogeneous and loosely coupled
- INFN Cloud core services and some centralized, fully managed, high level services are hosted on the Backbone
  - ✓ allows to leverage high-availability and disaster recovery capabilities
  - ✓ ensuring critical services always available



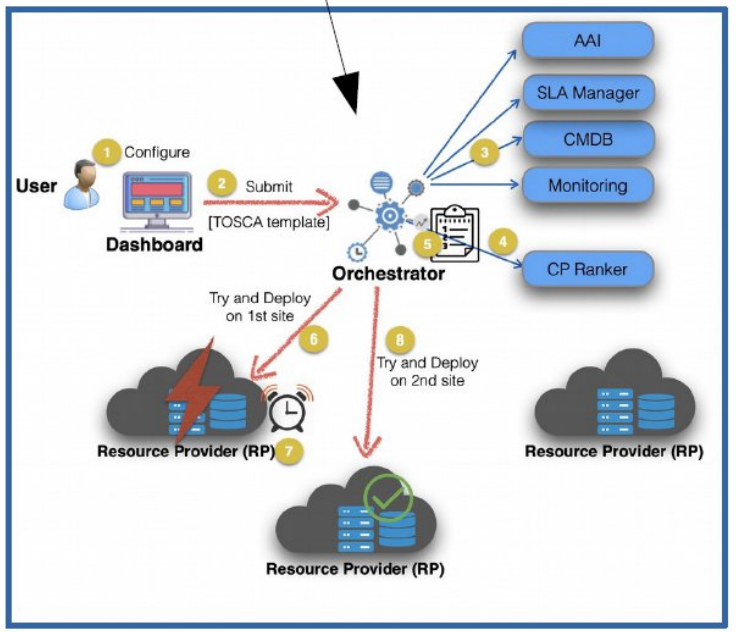
Federation middleware

Based on two outcomes from the INDIGO-DataCloud EU project:

- **INDIGO-IAM** (Identity and Access Management)
  - ✓ born to replace the VOMS
  - ✓ based on Oauth2 and OIDC standard protocols
  - ✓ support legacy AAI solutions
- **INDIGO-PaaS** Orchestrator
  - ✓ Enables the federation of distributed and heterogeneous compute environments:
    - Clouds
    - Container orchestration platforms
    - HPC systems
  - ✓ Smart scheduling based on compute/storage requirements vs provider capabilities
  - ✓ Client interfaces for advanced users (REST APIs, CLI python bindings) and end-users (web dashboard)



INDIGO DataCloud



Service portfolio

Fully managed centralized services (SaaS)

INFN Cloud object storage
the centrally managed service based on Ceph Rados-Gateway
GO TO SERVICE →

INFN Cloud Registry
The centrally managed service INFN Cloud Registry, based on Harbor
GO TO SERVICE →

INFN-Cloud monitoring
the INFN-Cloud monitoring service
GO TO SERVICE →

Notebooks as a Service (NaaS)
Jupyter Notebooks as a Service
Scope: admins/beta-testers
GO TO SERVICE →

On-demand self managed services (PaaS)

Virtual machine
Launch a compute node getting the IP and SSH credentials to access via ssh
CONFIGURE →

Docker compose
Deploy a virtual machine with docker engine and docker compose pre-installed.
Optionally run a docker compose file fetched from the specified URL.
CONFIGURE →

Run docker
Run a docker container
CONFIGURE →

INDIGO IAM as a Service
The on-demand deployment service for the INDIGO IAM provides a quick and easy way for organizations to deploy their own instance of the INDIGO IAM, which is an open-source Identity and Access...
CONFIGURE →

Elasticsearch and Kibana
Deploy a virtual machine pre-configured with the Elasticsearch search and analytics engine and with Kibana for simple visualization of data with charts and graphs in Elasticsearch
CONFIGURE →

Kubernetes cluster
Deploy a single master Kubernetes cluster
CONFIGURE →

Spark + Jupyter cluster
Deploy a complete Spark + Jupyter Notebook on top of a Kubernetes (K8s) computing cluster
CONFIGURE →

HTCondor mini
Deploy HTCondor mini, a technology preview of an all-in-one ("minicondor") HTCondor. This type of install is useful for testing and experimentation.
CONFIGURE →

HTCondor cluster
Deploy a complete HTCondor cluster
CONFIGURE →

Jupyter with persistence for Notebooks
Run Jupyter on a single VM enabling Notebooks persistence
CONFIGURE →

Jupyter + Matlab (with persistence for...
Run Jupyter on a single VM enabling Notebooks persistence and Matlab integration
CONFIGURE →

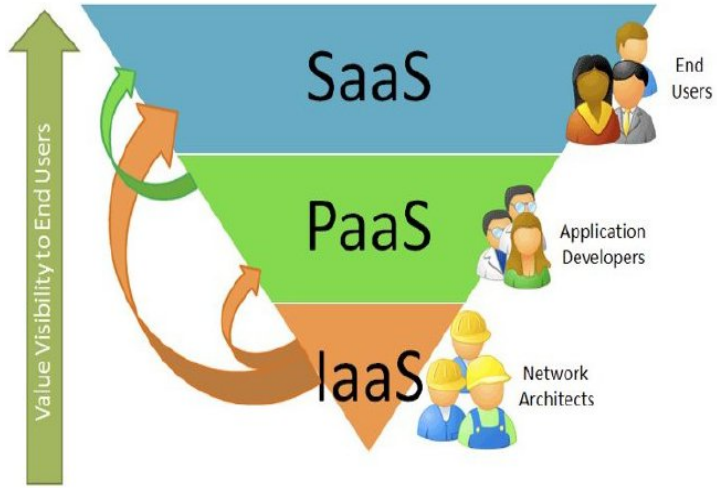
Computational enviroment for Machine Le...
Run a single VM with exposing both ssh access and multiuser JupyterHub interface, integrating the AI-INFN environment
CONFIGURE →

Working Station for CYGNO experiment
Run a single VM with all the CYGNO enviroment exposing both ssh access and Jupyter
CONFIGURE →

Sync&Share aaS
The INFN-Cloud Sync&Share aaS is based on popular storage solutions suche as ownCloud [1] and Nextcloud [2].
INFN-Cloud users have full control over the configuration parameters ...
CONFIGURE →

Infrastructure services (IaaS)

- Start, Stop, Delete a VM
- Hostname choice
- Manage VM ports



Service implementation strategy

- Based on the Infrastructure as Code (IaC) paradigm
- Describe at high level “What” is needed rather than “How” a service or a functionality should be implemented
- Key technologies:

TOSCA is used to model the topology of the whole application stack

Ansible is used to automate the configuration of the virtual environments

Docker is used to encapsulate the high-level application software and runtime

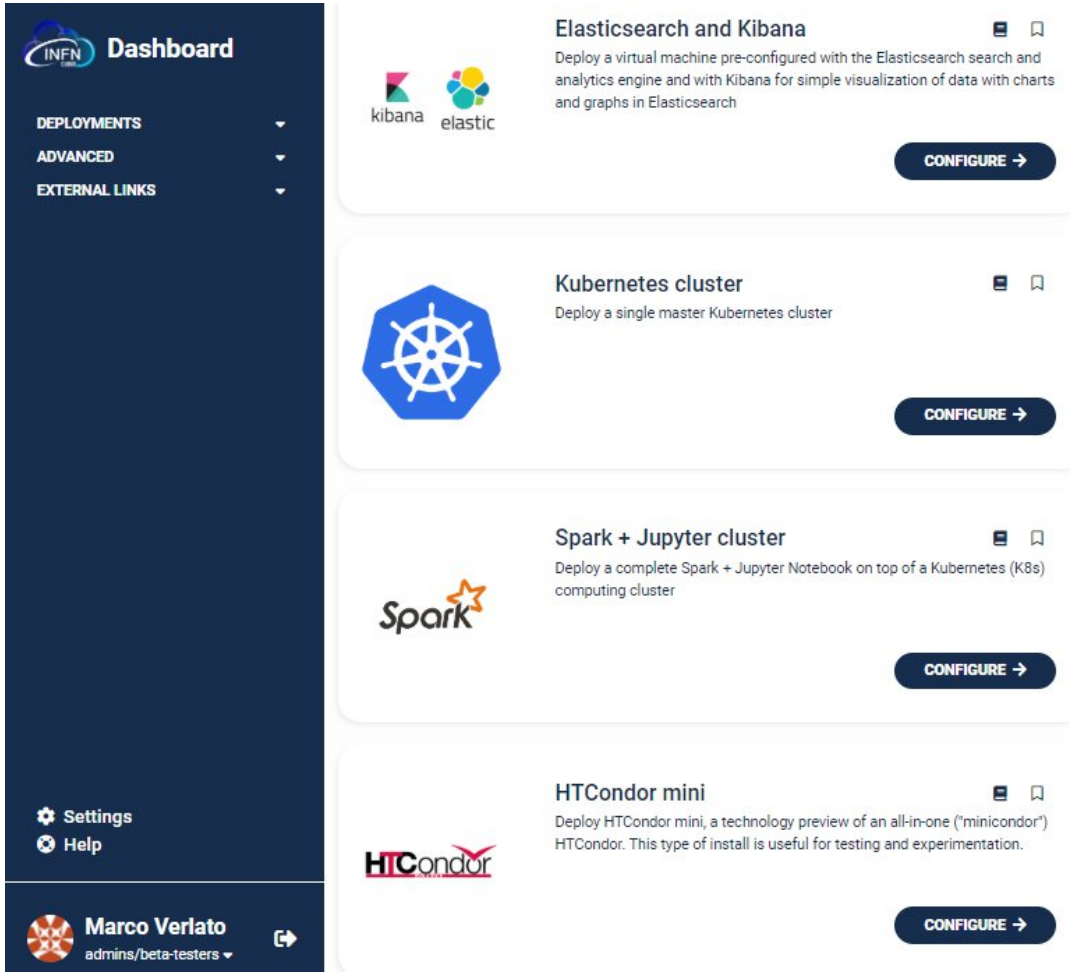
- Enabling a Lego-like approach: services can be composed and built on top of re-usable components to create the desired infrastructure

User front-end

INDIGO-PaaS Dashboard

https://my.cloud.infn.it

- Web-based intuitive user interface
- Enables users to manage and monitor their deployments
- No TOSCA knowledge required
- Hides all technical details
- OIDC authentication
- Multi-tenancy
- Secret management (via Vault integration)
- Dynamic view of service portfolio (depends on use group membership)



Self-provisioning

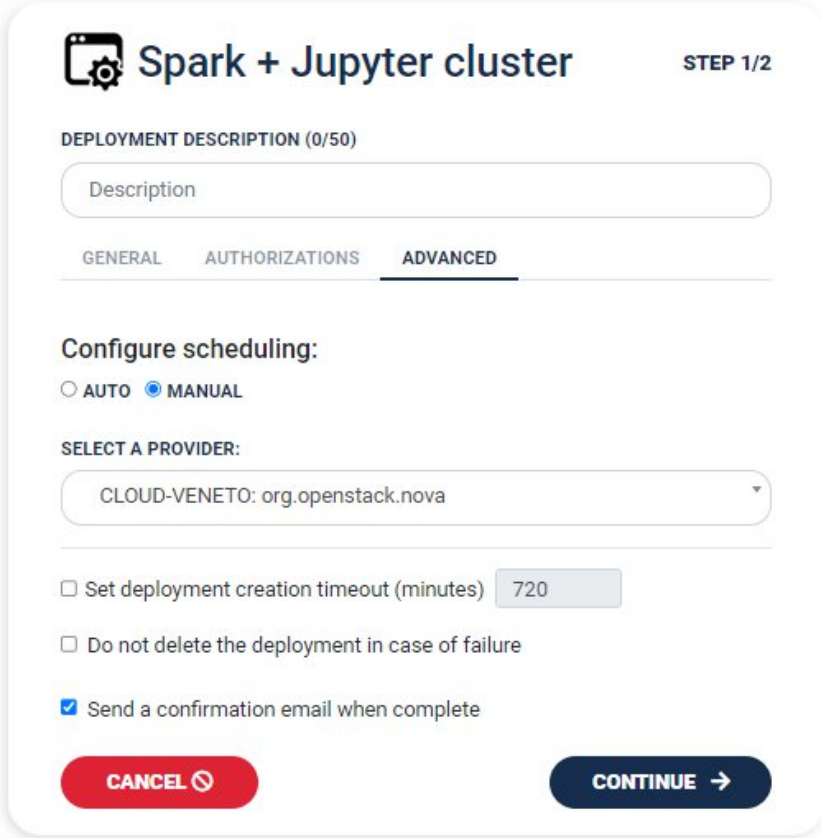
Requesting a service with just few clicks

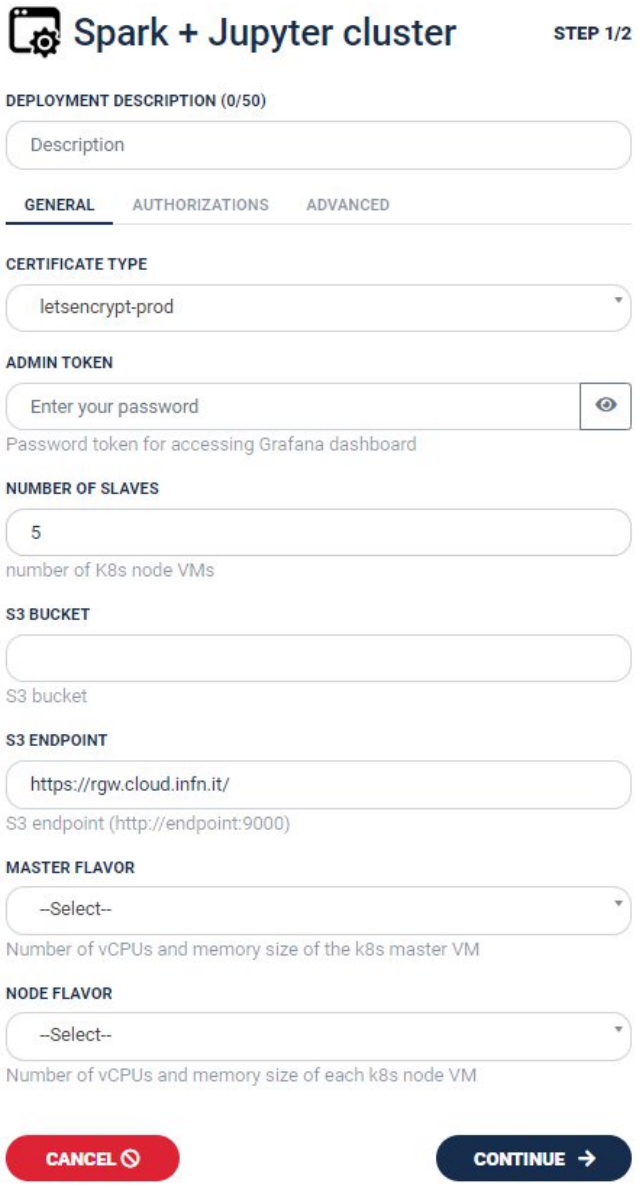
- Insert input parameters to customize the deployment

- Choose the scheduling strategy:

  ✓automatic: let the Orchestrator to select the best provider

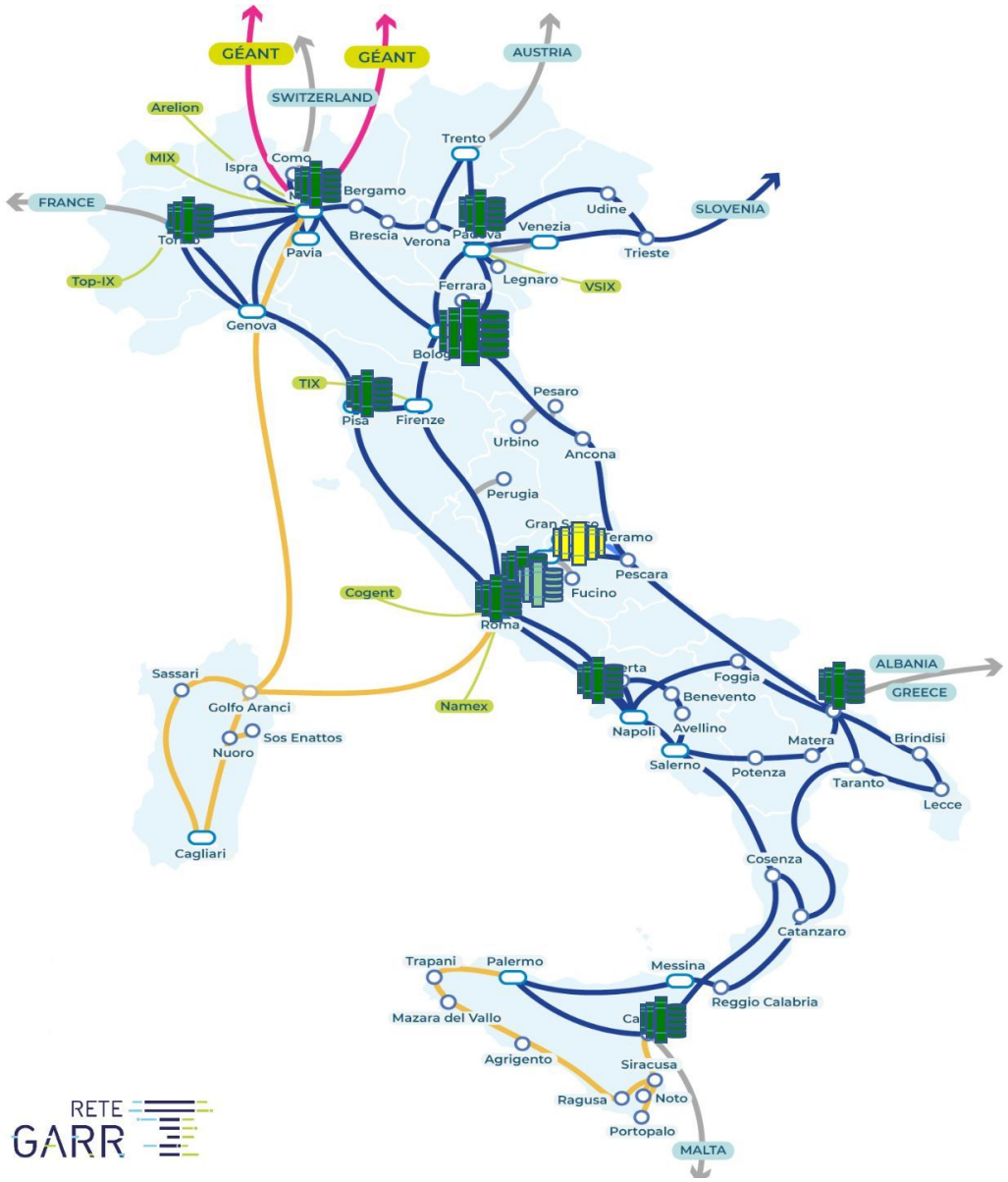
  ✓manual: choose the provider from the list returned by the SLA Manager service





DataCloud

- INFN Cloud is part of a wider infrastructure for the INFN Scientific Computing
  - ✓ Tier-1 (CNAF)
  - ✓ Tier-2's (BA, CT, LNF, LNL/PD, NA, MI, PI, RM1, TO)
  - ✓ INFN Cloud (Backbone and federated clouds)
  - ✓ HPC4DR (LNGS)
  - ✓ (Tier-3)
- DataCloud is evolving into a Cloud Federation, following the INFN Cloud model
  - ✓ resources made available through cloud interfaces
  - ✓ easy of use, through the PaaS Orchestrator and dashboard
- Traditional (Grid and batch system) access remains as needed and when convenient
  - ✓ e.g. through VK offloading



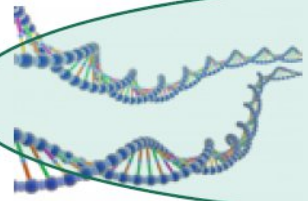
Increasing computing requirements

Il Large Hadron Collider (LHC) al CERN

1 LHC Experiment ~2020: ~200.000 CPU Cores; ~200 PB disk; ~350 PB tape → 1 High-Luminosity LHC Experiment ~2028: ~15M CPU Cores; ~15 EB disk; ~26 EB tape

DUNE: ProtoDUNE in 2019 collected 3 GB/s; real DUNE expected 80x at the end of the 2020s.

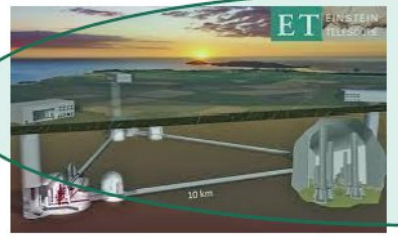
SKA: up to 2 PB/day , to be collected and processed at "complex" locations.



Genomics: a single genome ~100 GB. Any population study (>1M people) over 100 PB

CTA: ~ 10 PB/y in 2025+.

Virgo: ~10% of a LHC experiment.



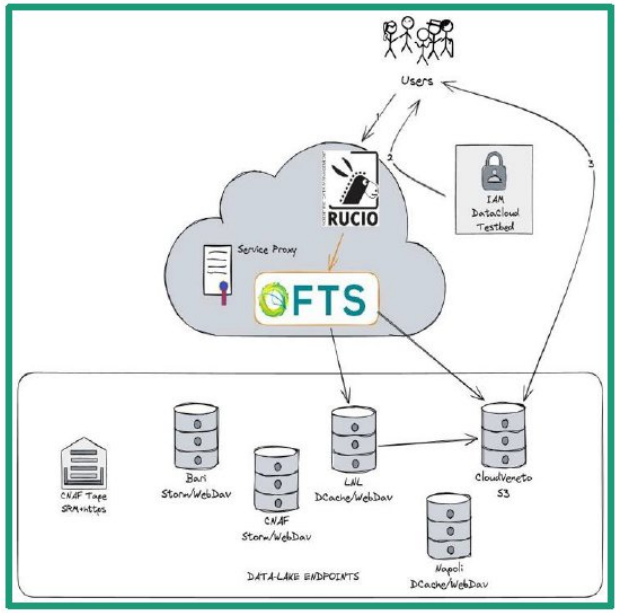
Einstein Telescope: aiming at being ~10% of a HL-LHC experiment.

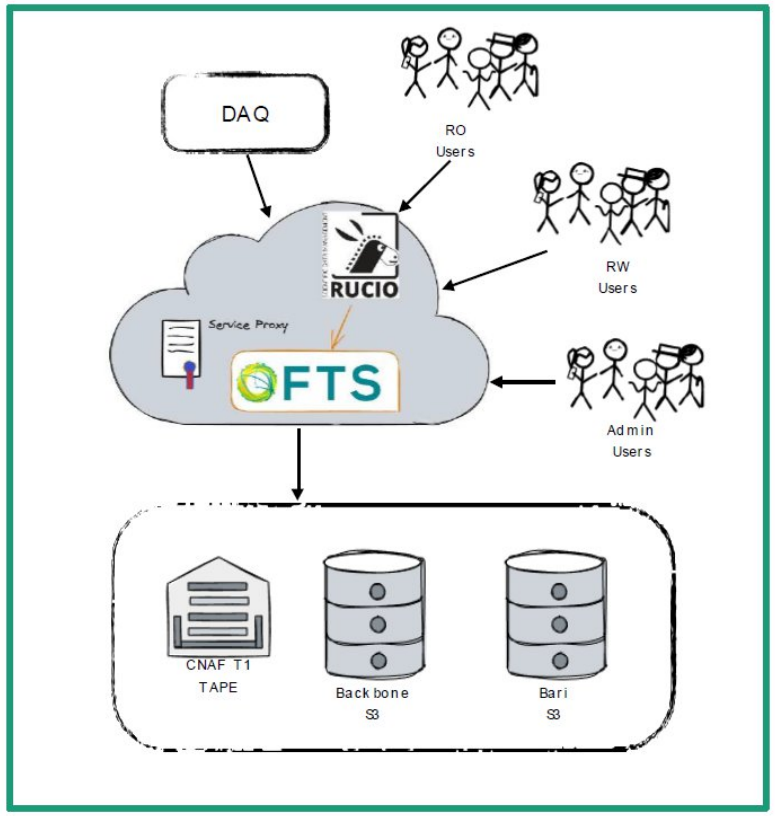
- DataCloud addresses the needs of INFN research projects
- Computing requirements of no HEP experiments are increasing too
- INFN is revisioning and expanding its infrastructure and services adopting a "cloud first" approach

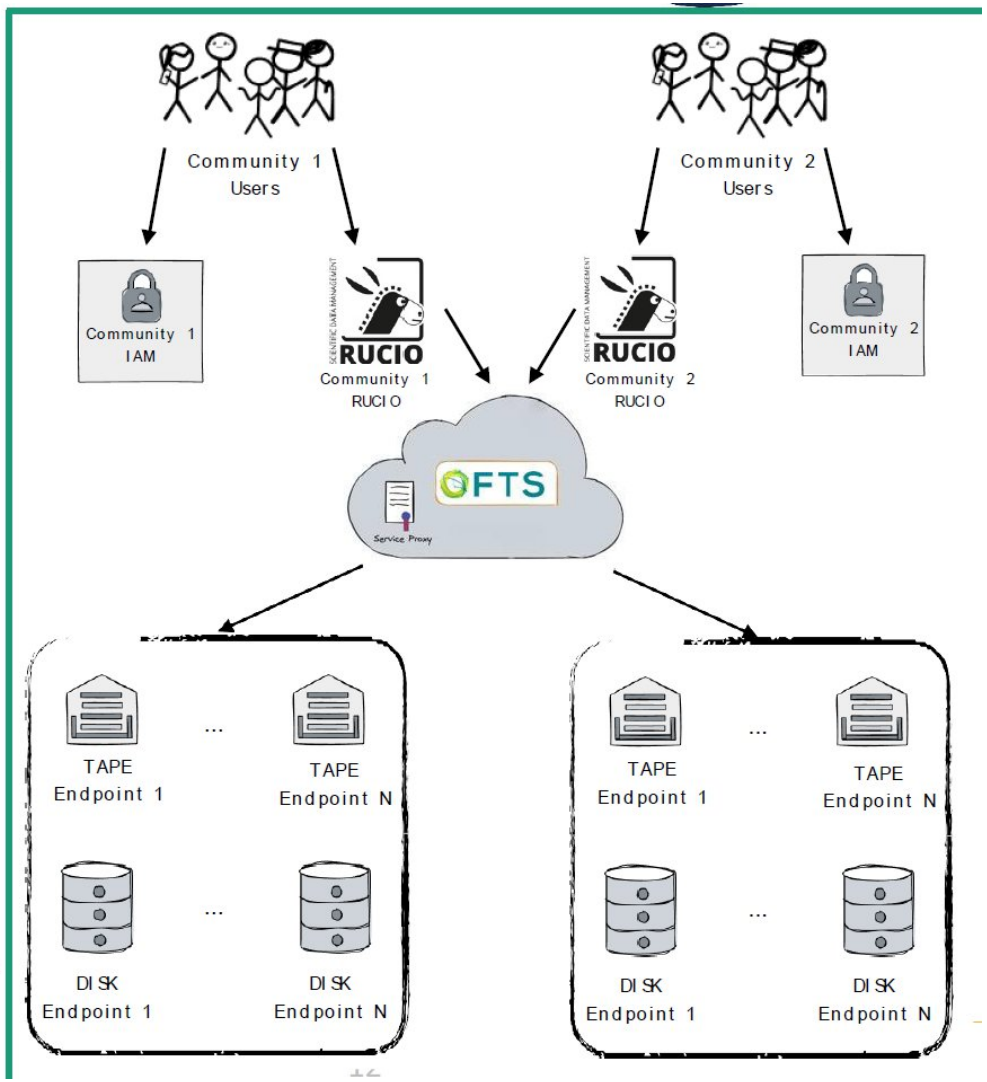
The “Cloud-Data lake” model

- It is the evolution of the computing infrastructure for both HEP and not HEP experiments
- Resources no longer provisioned only through dedicated grid sites
  - ✓inclusion of HPC systems and commercial clouds as part of the resources we must be able to take advantage of
    - INFN already demonstrated the capability to execute LHC workflows on HPC systems (in particular @CINECA) and on commercial clouds (e.g. ARUBA).
- Optimize storage access and management
  - ✓reduce the number of replicas, few big sites high speed connected
  - ✓CPU and storage no longer coupled together
  - ✓deploy caches where needed

Data lake implementation







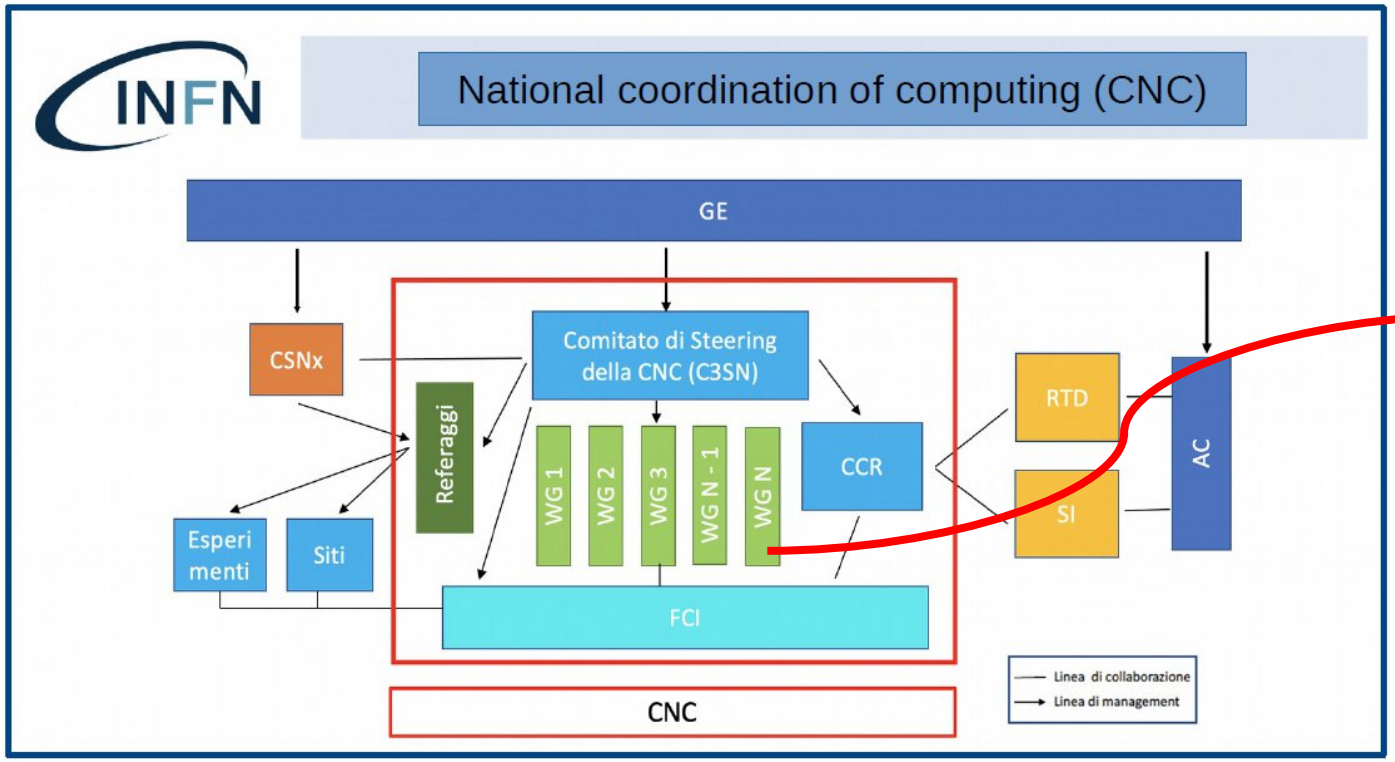
Key components:

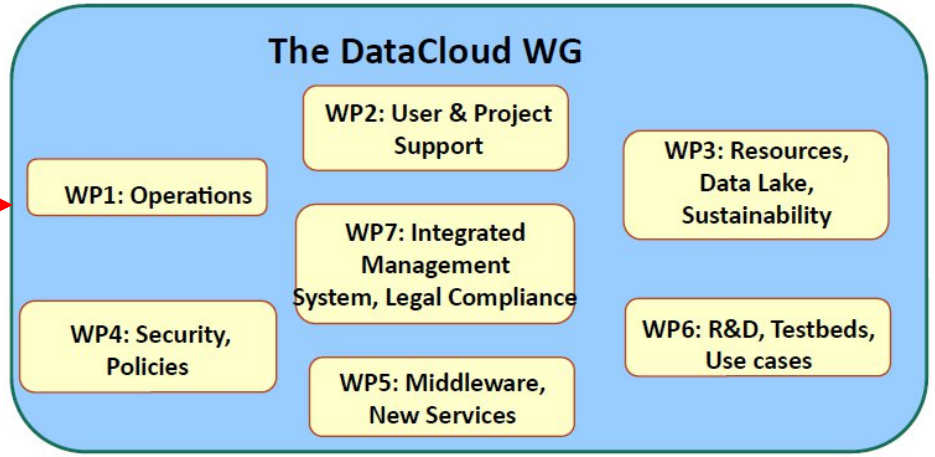
RUCIO
SCIENTIFIC DATA MANAGEMENT

FTS
File Transfer Service

Governance

DataCloud is also a Working Group of the INFN CNC, responsible to manage and evolve the INFN distributed infrastructure and services



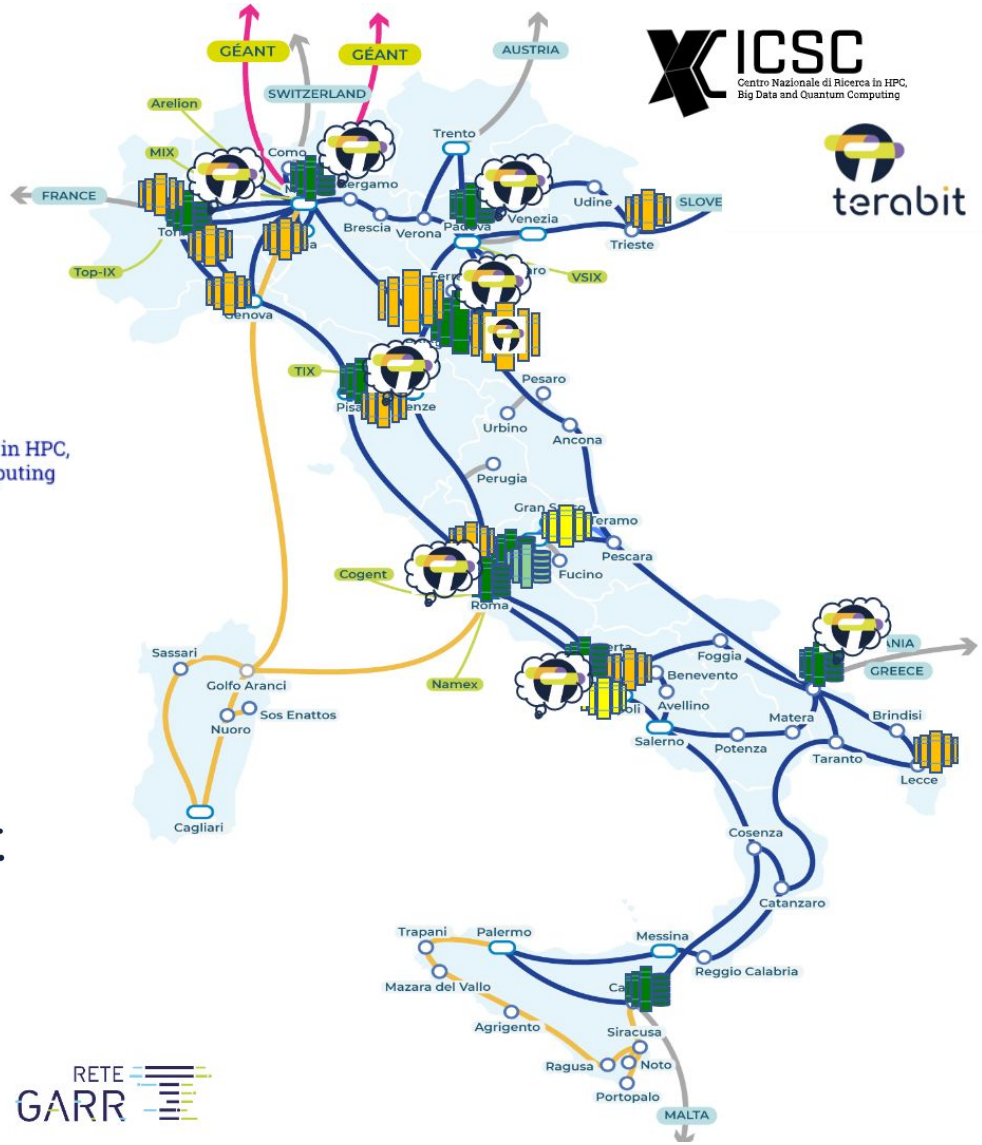


DataCloud and NRRP projects

INFN has a central role in the Italian National Recovery and Resilience Plan (NRRP) computing related initiatives

- **ICSC - National Research Centre in HPC, Big Data and Quantum Computing**
  - ✓ 10 thematic spokes
  - ✓ 1 infrastructure spoke (INFN, CINECA, GARR)
  - ✓ 25 universities, 12 research institutes, 14 strategic private companies
  - ✓ 320 M€ budget

- **TeRABIT - Terabit network for Research and Academic Big Data in ITaly**
  - ✓ Partners are the same of the ICSC infrastructure spoke (INFN, CINECA, GARR)
  - ✓ Covers areas complementary to those of the ICSC infrastructure
  - ✓ 41 M€ budget

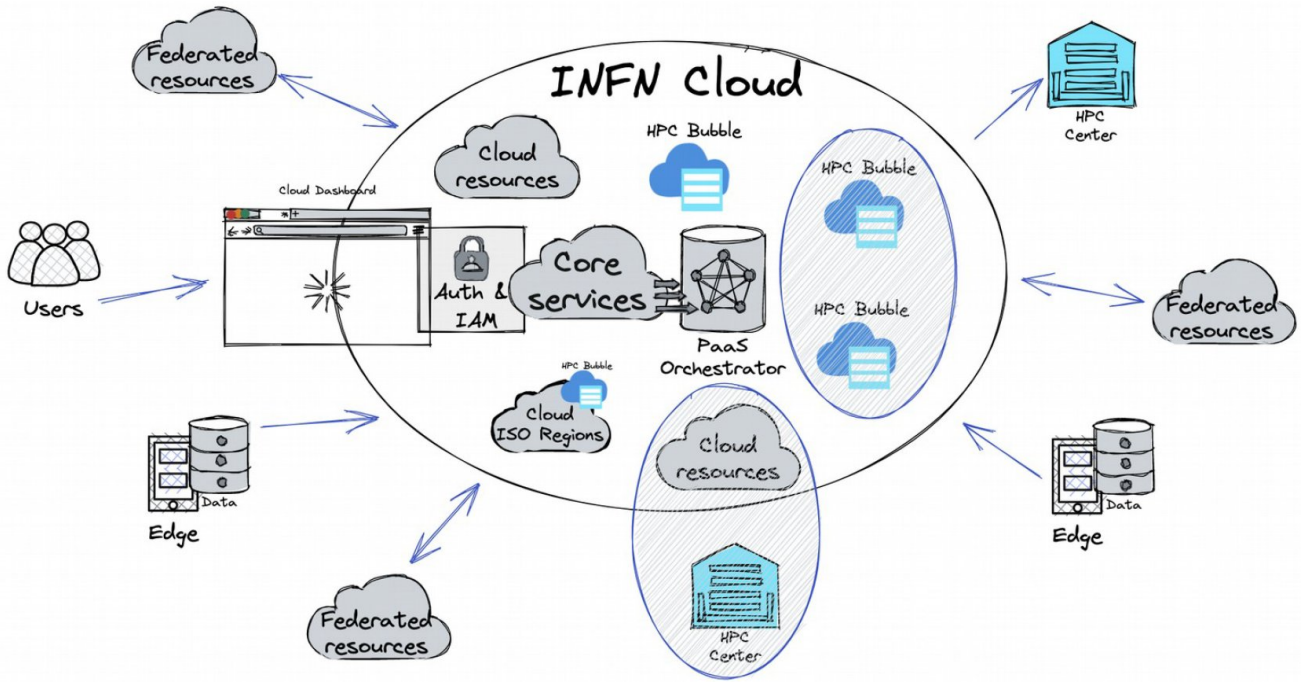


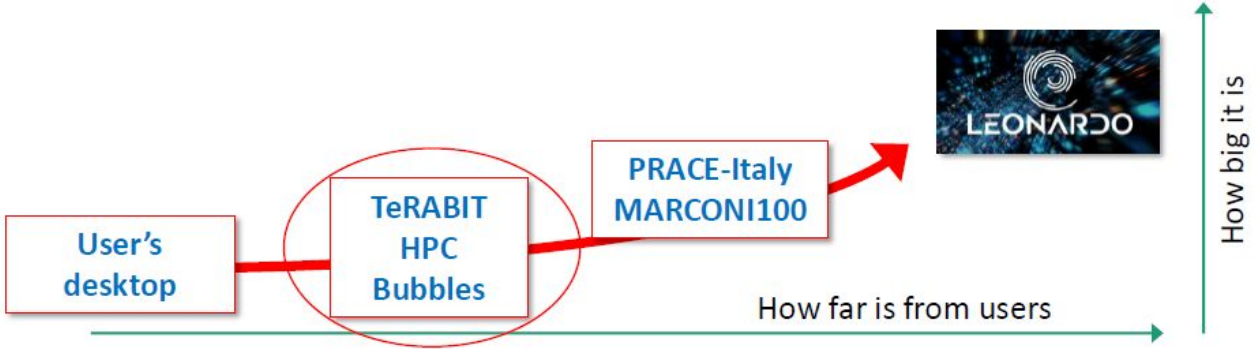
DataCloud for the Italian Cloud Federation

In particular, the TeRABIT project envisions the creation of a a distributed, hyper-connected, hybrid HPC-Cloud environment:

- integrating the distributed INFN infrastructure with HPC resources of PRACE-Italy (CINECA), through a high speed network provided by GARR and complementing the ICSC National Center
- enabling widespread data transfer, up to Terabits per second, and services on a national scale in Italy, connected to Europe
- innovating the central HPC node of PRACE-Italy
- innovating the HPC services offered to researchers, beyond the centralized calculation model, adding distributed “HPC-Bubbles”
  - ✓ Clusters with CPUs, CPUs+GPUs, CPUs+FGPAs
  - ✓ Infiniband network and fast storage
  - ✓ e.g. at INFN-PD 4500+ HT Cores Bubble

The aim is to realize a scalable open “Edge-Cloud Continuum”





Conclusions

- After 10 years of activity, CloudVeneto is a successful example of fruitful collaboration between the INFN and UniPD Departments
  - ✓ it is a production service, well supported, used with satisfaction by users and continuously expanding
  - ✓ it allows a more efficient and flexible use of processing and storage resources
  - ✓ It has been the first cloud site outside the backbone joining the INFN Cloud infrastructure
- Due to growing needs and substantial new opportunities, INFN is expanding its computing infrastructure and services
  - ✓ its approach is to abstract from where resources are located, leveraging aaS models to build a cloud-native, national, federated structure
  - ✓ with the ambition to create, manage and develop a vendor-neutral, open, scalable and flexible “data lake” that serves much more than just INFN users and experiments
  - ✓ to become a key asset for fundamental, applied and industrial research in Italy and beyond

Thanks for your attention

Questions?

The CloudVeneto Team

INFN-Padova

Paolo Andreetto
Alberto Crescente
Federica Fanzago
Rita Lenzo
Massimo Sgaravatto
Sergio Traldi
Antonino Troja
Marco Verlato
Lisa Zangrando

INFN-LNL

Sergio Fantinel

UniPD

Loris Lazzaro
Paolo Emilio Mazzon
Matteo Menguzzato
Giampietro Sella