



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani

PIANO NAZIONALE
DI RIPRESA E RESILIENZA



Centro Nazionale di Ricerca in HPC,
Big Data and Quantum Computing

The Scientific Hub as a tool for astrophysics and cosmology communities

*Matteo Pasqui
(INFN Sezione di Tor Vergata)*

Spoke 3 II Technical Workshop, Bologna Dec 17 -19, 2024

Where were we

Elba meeting

We get started then:

- finding volunteer use cases
- deploying the Hub on "in-kind" resources from INFN
 - waiting for official resources to come

It might be technically possible to reproduce the interTwin PoC meaning to "offload" from an hub hosted on the cloud to HPC resources

Interactive prototyping

Successfully spawning on-demand jupyterLab instance on HPC along with other more "cloud-ish" instance on K8s.

Crucial for experimenting ML workflows interactively before scaling out!

Access to an extensible HUB

- o Hosting core services on cloud resource
 - o Extend the resource pool for interactive sessions over HPC and, more generally, distributed resources
- There are then cases where automation and scalability comes without (or after) interactive development



Our pillars

WP5 should act as a two-ways bridge towards (and from) Spoke0:

- Bringing well defined use cases as a validation tool for the infrastructure PoC
- Getting access to early development decisions

Where are we now (SPOILER ALERT!):

- The HUB implementation for Spoke3 is part of the DataCloud - CINECA integration PoC activity
 - effectively one of the first evaluator to get access to the offloading solution (along with Spoke2)
- Crucial to get a significative sample of experiences to drive the development of new features accordingly to our needs

Scientific Rationale

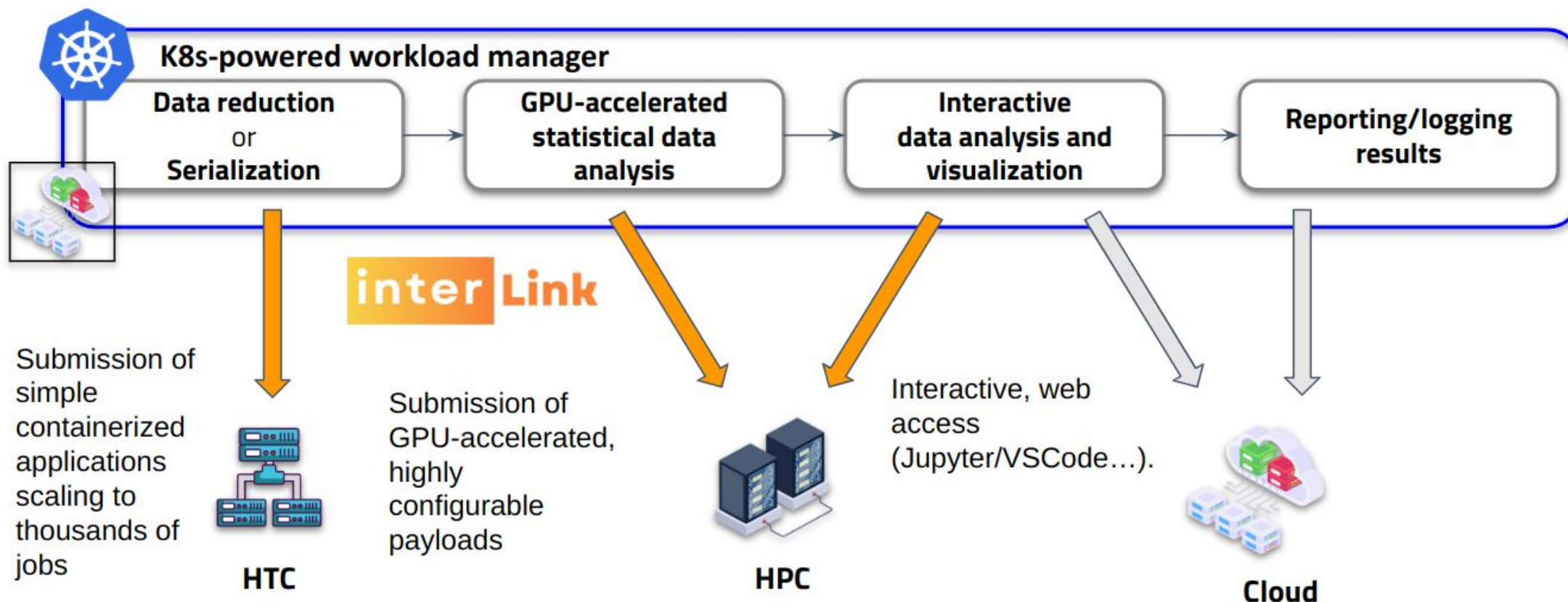
Modern astrophysics and cosmology demand massive computational resources for data processing and simulations.

Challenges:

- Fragmented access to HPC and cloud resources
- Inefficient task allocation and resource utilisation
- High complexity in managing multi-stage workflows

Solution: A Scientific Hub that integrates HPC and cloud resources, providing a unified, scalable, and user-friendly platform.

Offloading each task to the proper resource kind



Synergies with EU projects ([interTwin](#)) and INFN initiatives (e.g. this diagram is courtesy of M.Barbetti and L. Anderlini talk <https://indi.to/gqWcc>)

Use cases

- ***Machine Learning***: Accelerate training and inference on astrophysical datasets
- ***Astrophysics and Cosmology Simulations*** (e.g. PINOCCHIO: simulation of dark matter halos)
- ***Real-Time Resource Management***: Predictive algorithms for dynamic task allocation

Use case: PINOCCHIO

PINOCCHIO simulates dark matter halo formation using Lagrangian Perturbation Theory (LPT).

Hub's role:

- Access to HPC systems, such as the Leonardo supercomputer, for large-scale simulations
- Dynamically allocation of resources to optimise simulation time
- Seamless collaboration through sharing workflows and results

Impact:

- Enhanced precision in cosmological simulations
- Accelerated processing time for large-scale computations

Use case: Cosmic Ray Background Data Generation

The Hub is a valid tool to employ in scenarios that concern large-scale data generation.

How the Hub enhances the process:

- HPC/HTC integration for the computationally expensive generation of data
- seamless execution and training of Generative Adversarial Networks (GANs) and other machine learning workflows

(e.g. GANs for cosmic background simulation for the LiteBIRD experiment, G.Cavallotto)

Our strategy for data handling

Enable and enhance the interoperability between heterogeneous storage solutions in the ICSC distributed infrastructure

- big traditional HPC centres, Cloud

Key technology



Strict collaboration with Spoke0 related activities

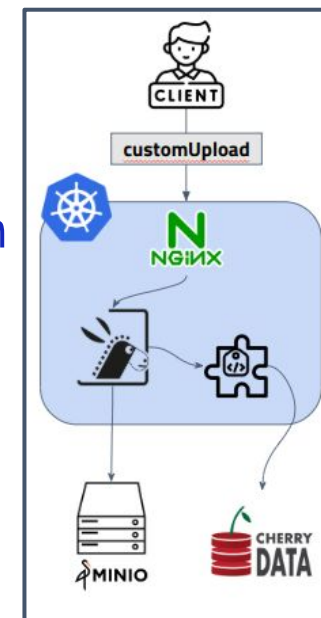
To promote the adoption of RUCIO based ecosystem in order to grant to the user "The ability to store and retrieve data without caring about to know where the data actually is"

Synergies with Innovation Grant

Enhancing RUCIO ecosystem.

A ongoing synergy with **interoperable Data Lake (IDL) project**

example of Spoke 3-2-0 cross activity



(see Diego's WP4 talk)

We are ready and already testing

We have a system online where the presented use case have access to a web interface for interactive analysis

Users can login and start developing based on their needs.

All sitting "seamlessly" on a Leonardo@CINECA node.

N.B. the IAM system from ICSC is coming soon, we need to follow up for AuthN/Z integration at all levels of the system

N.B. in-kind resources, ICSC migration coming soon
<https://jhub.131.154.98.40.myip.cloud.infn.it/>

Server Options

Select your desired image:

Select your desired number of cores:

Select your desired memory size:

Enable Offloading to No
 Leonardo CPU
 Leonardo GPU

User ▲	Admin ▼
<input type="button" value="Add Users"/>	
dciangot	admin
fcuna	
gcavallotto	
mbossa	

Main Results

- Tests are being done right now!
- Test by Federica Cuna on Leonardo GPUs shows good parallelisation

Terminal interactive visualisation of GPU utilization (in percent values) and memory usage

```

+-----+-----+-----+-----+-----+-----+-----+-----+
| NVIDIA-SMI 530.30.02                Driver Version: 530.30.02    CUDA Version: 12.1    |
+-----+-----+-----+-----+-----+-----+-----+-----+
| GPU  Name                Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf          Pwr:Usage/Cap|      Memory-Usage | GPU-Util  Compute M. |
|=====+=====+=====+=====+=====+=====+=====+=====+
|   0   NVIDIA A100-SXM-64GB         On          | 00000000:1D:00:0 Off |              0      |
| N/A   44C    P0              83W / 459W | 7771MiB / 65536MiB |    1%      Default  |
|                                           |                      | Disabled   |
+-----+-----+-----+-----+-----+-----+-----+-----+
|   1   NVIDIA A100-SXM-64GB         On          | 00000000:56:00:0 Off |              0      |
| N/A   45C    P0              86W / 461W | 3999MiB / 65536MiB |   70%      Default  |
|                                           |                      | Disabled   |
+-----+-----+-----+-----+-----+-----+-----+-----+
|   2   NVIDIA A100-SXM-64GB         On          | 00000000:8F:00:0 Off |              0      |
| N/A   45C    P0              88W / 458W | 3957MiB / 65536MiB |   86%      Default  |
|                                           |                      | Disabled   |
+-----+-----+-----+-----+-----+-----+-----+-----+
|   3   NVIDIA A100-SXM-64GB         On          | 00000000:C8:00:0 Off |              0      |
| N/A   44C    P0              82W / 457W | 4103MiB / 65536MiB |   10%      Default  |
|                                           |                      | Disabled   |
+-----+-----+-----+-----+-----+-----+-----+-----+
| Processes:                                |
| GPU  GI  CI           PID  Type  Process name                        GPU Memory |
|   ID  ID  ID             |          |          |                                     Usage      |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

Final Steps

- The Scientific Hub simplifies and optimises the use of CINECA and Cloud resource in synergy with Spoke0 activities
- By leveraging a Kubernetes-based architecture, it ensures *interoperability and reproducibility of the whole system*
- Validating frontiers use cases in astrophysics and cosmology gives us the capability to participate to the integration activity in a “proactive” fashion

Future work will focus on:

- reporting back to Spoke0 and CINECA our results in the context of the integration PoC
- migration to ICSC IAM system for AuthN/Z
- migration to ICSC provided CLOUD resources



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



THANK YOU