



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



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

WP5: Architectural Support for Theoretical and Experimental Physics Data Management on the Distributed CN infrastructure

WP5 Leader: Daniele Spiga (INFN – sez. Perugia) & Elvira Rossi (Università Federico II di Napoli)

Agenda of the Talk

- Recap of WP5 mission
- Overview of the main achievements and current status
- Roadmap for the upcoming months
- Summary and Lessons after 12 months

WP5 Objectives and high level view

Proposes:

- 5.1 Support of the adaptation of existing applications on the data-lake distributed infrastructure, and via innovative computational models
- 5.2 Competence center for the design, implementation and test of computing models

Objectives:

- ✓ O5.1: document and report best practices for integrations with the CN datalake
- ✓ O5.2 prepare tools to ease integration with the CN infrastructure
- ✓ O5.3 Offer support for transitioning the computing models
- ✓ O5.4 organize training opportunities open to external users

WP5		
	Nome	Mail
National coordinators	Elvira Rossi Daniele Spiga	elvira.rossi@unina.it daniele.spiga@pg.infn.it
Institute coordinators		
INFN	Daniele Spiga	daniele.spiga@pg.infn.it
UNIMIB	Mattia Bruno	Mattia.Bruno@mib.infn.it
UNINA	Elvira Rossi	elvira.rossi@unina.it
UNITS	Andrea Bressan	andrea.bressan@ts.infn.it
UNIBO	Alessandra Fanfani	Alessandra.fanfani2@unibo.it
UNIFE	Luca Tomassetti	luca.tomassetti@unife.it

WP5: indico pages & email list and meeting

Spoke 2: <https://agenda.infn.it/category/1774/>

FUNDAMENTAL RESEARCH & SPACE ECONOMY

General	5 events
WP1	10 events
WP2	20 events
WP3	12 events
WP4	15 events
WP5	19 events
WP6	9 events
Steering	19 events
General Assembly	7 events
Discussioni Industrie	8 events
Varie	1 event
Innovation Grants	9 events

Established a bi-weekly meeting to discuss progresses and activities to:

- keep everybody up to date
- gather requirements from users
- solve problems.

WP5: <https://agenda.infn.it/category/1781/>

December 2023	20 Dec	Biweekly Meeting Spoke2 - WP2.5
	13 Dec	Biweekly Meeting Spoke2 - WP2.5
November 2023	22 Nov	Biweekly Meeting Spoke2 - WP2.5
	10 Nov	Biweekly Meeting Spoke2 - WP2.5
	08 Nov	WP5 - Analysis Facility
	03 Nov	Biweekly Meeting Spoke2 - WP2.5
October 2023	13 Oct	Biweekly Meeting Spoke2 - WP2.5
July 2023	07 Jul	Biweekly Meeting Spoke2 - WP2.5
June 2023	09 Jun	Spoke 2 - WP5 & WP2 - ATLAS
May 2023	24 May	Biweekly Meeting Spoke2 - WP2.5
April 2023	28 Apr	Biweekly Meeting Spoke2 - WP2.5
	12 Apr	Biweekly Meeting Spoke2 - WP2.5
March 2023	22 Mar	Biweekly Meeting Spoke2 - WP2.5
	09 Mar	Meeting Spoke2 - WP2.5
February 2023	02 Feb	Operational Spoke2 - WP2.5 Meeting: Perugia-Napoli
December 2022	15 Dec	Meeting Spoke2 - WP2.5
November 2022	04 Nov	Joint Meeting WP4 & WP5 Leaders
September 2022	29 Sept	Meeting Spoke2 - WP2.5
	23 Sept	Meeting Spoke2 - WP2.5

19 events: general WP5, operational and work meetings with WP5 contacts and WP5 enthusiast

Monthly/Biweekly/Weekly meetings as needed

Email list

Please contact us if you are interested in contribute to WP5

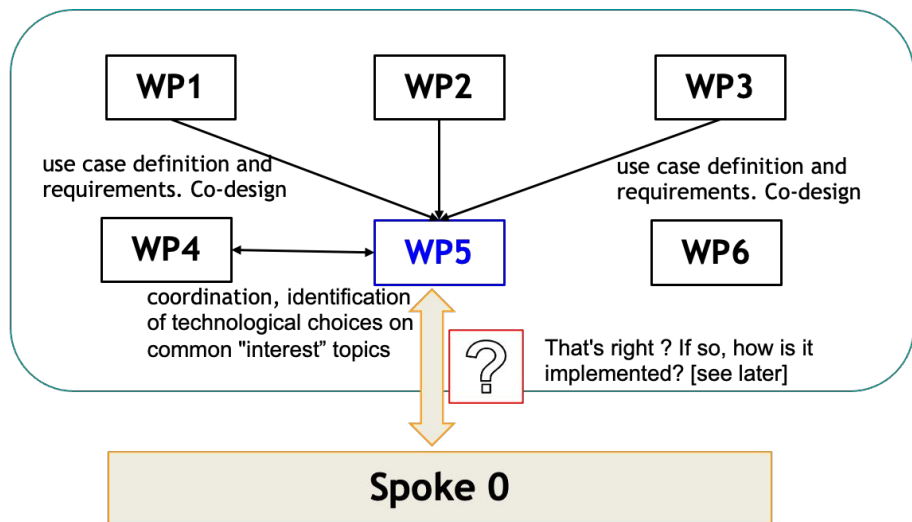
email list: cn1-spoke2-wp5-all@lists.infn.it

Link to subscribe to the email list:

<https://lists.infn.it/sympa/info/cn1-spoke2-wp5-all>

Positioning of WP5 in the global Spoke2 organization

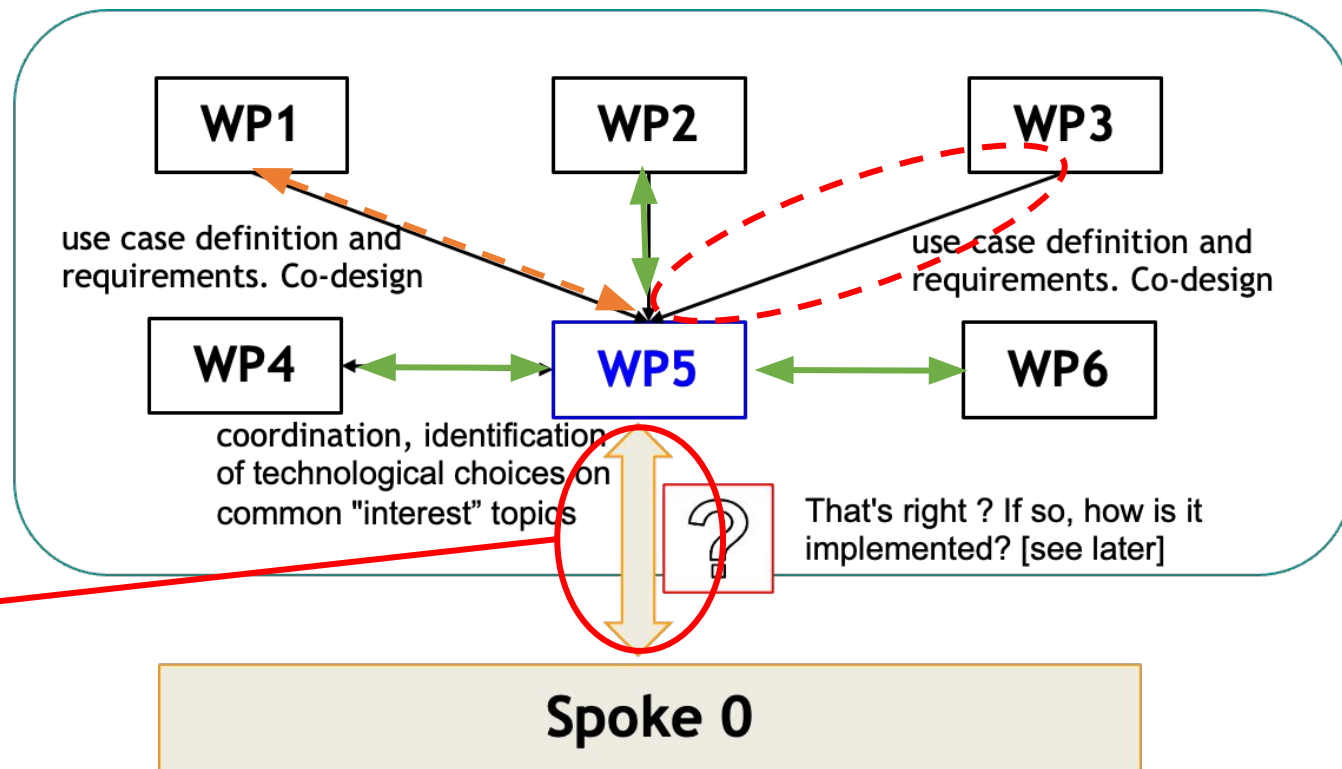
Our understanding/expectations one year ago



What we presented during the [kick-off](#)

Not yet implemented. However we have a more clear ideas on how it could work.
[see later]

What has been clarified/established after one year of activity



First year: Achievements in a nutshell

Together with WP4-6 we submitted a comprehensive survey to gather feedback and requirements

Best practices for data lake solutions and technologies for high rate analysis activities

- [Introduction](#)
- [Data Management and Data Access: needs and projections](#)
- [Data Lakes, Data Access and Processing](#)
- [Data Management and storage federation: Technologies](#)
- [High Rate Analysis: Technologies](#)
- [Summary and relationship with flagship use cases](#)

Document (O 5.1)

- High Rate Data Analysis

- a solution to use possibly a heterogeneous set of resources (seen from the user perspective: to transparently exploit different type of providers.... HPC, Cloud HTPC)
- a solution to access large amount of resources ALSO for (quasi)-interactive (to process a huge amount of data)

Prepare tools to easy exploit CN infrastructure (O 5.4)

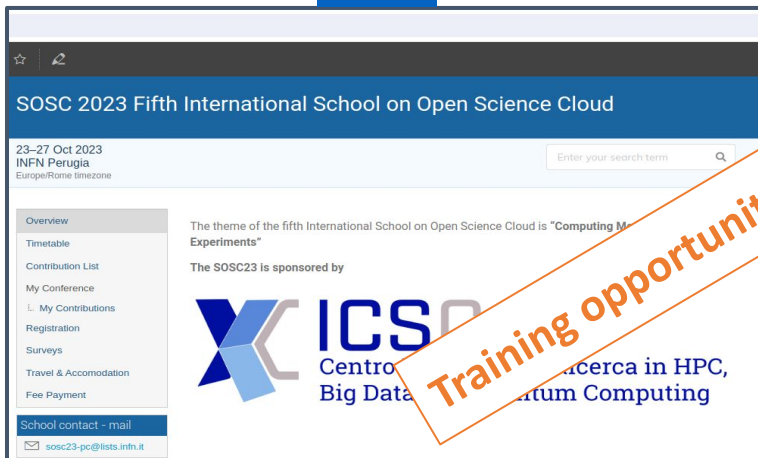
- Data Management in the Data Lake

- a solution to abstract the physical layer of the storage (not need to care about different backend)
- track data location
- automatically replicate data between different locations in a transparent manner (third party copy of data)

Preliminary: One testbed activated [see later]

Contributed to 2 Flagship preparation
See also: [F. Gravili Talk](#)

SOSC URL

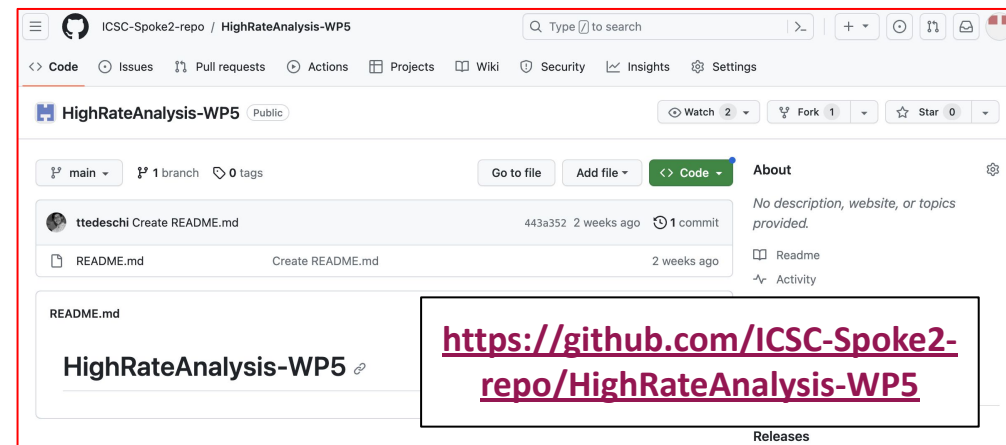


Training opportunities (O 5.4)

WP5 in practice: the implemented strategy

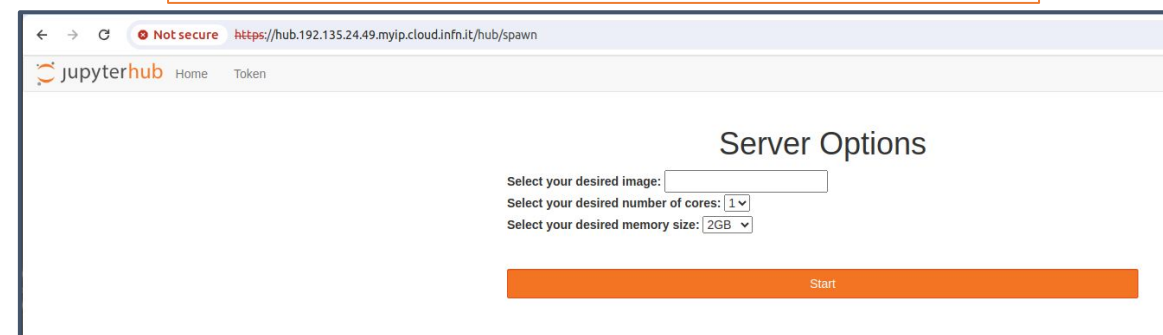
- To **build a community** of sw/computing experts and enthusiasts
- To **enable co-design** of (technical) solutions with sciences
- To **contribute to the overall CN infrastructure**
 - via Spoke 0 for infrastructure
 - open to inter-spoke activities

Sharing code



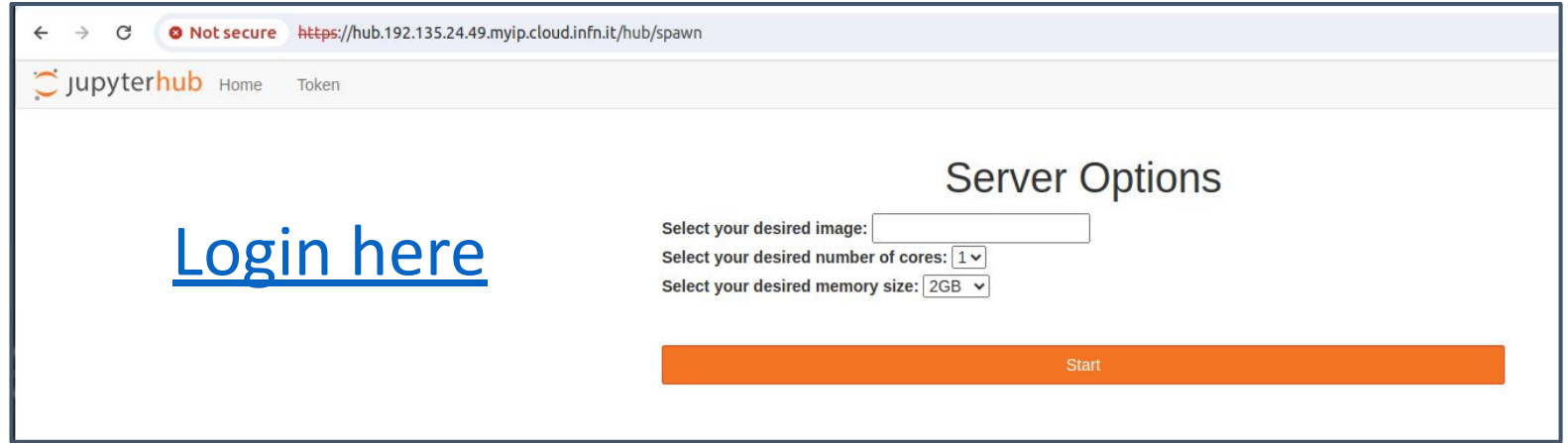
<https://github.com/ICSC-Spoke2-repo/HighRateAnalysis-WP5>

Deploying testbeds and playgrounds

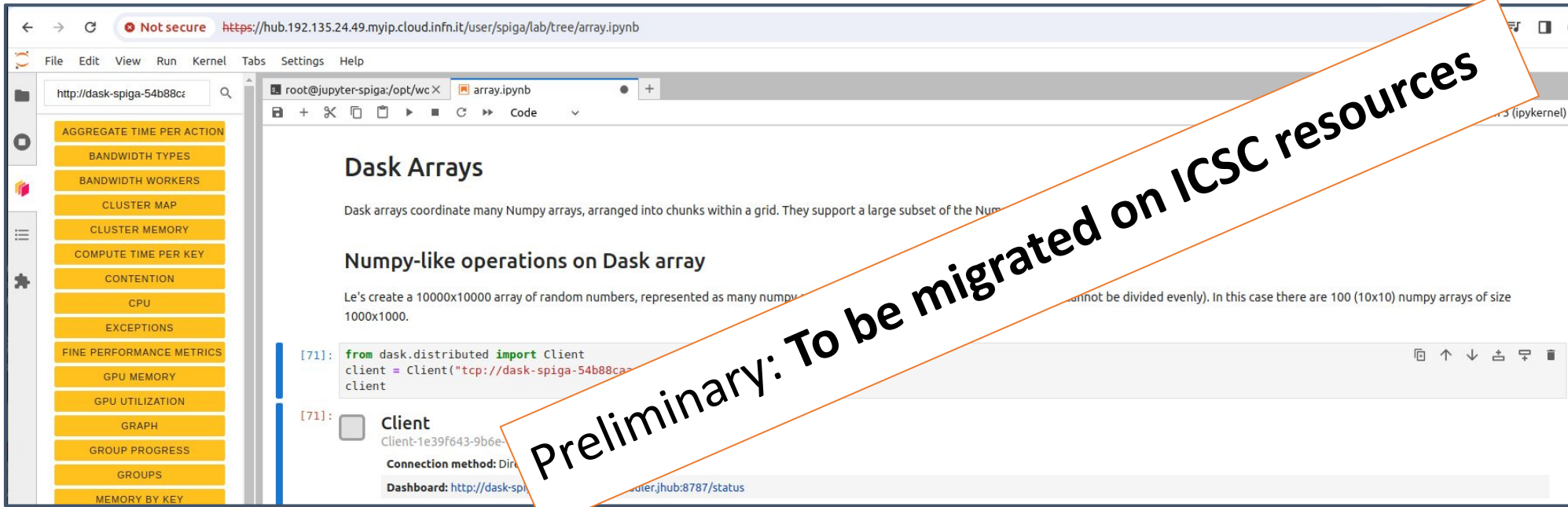


Our first testbed

We decided to try to “anticipate” the ICSC resource provisioning to start supporting WP1-2-3 (and to prepare flagships)



Temporary using a very small allocation on INFN-Cloud to prepare the SW configuration for the High Throughput data analysis

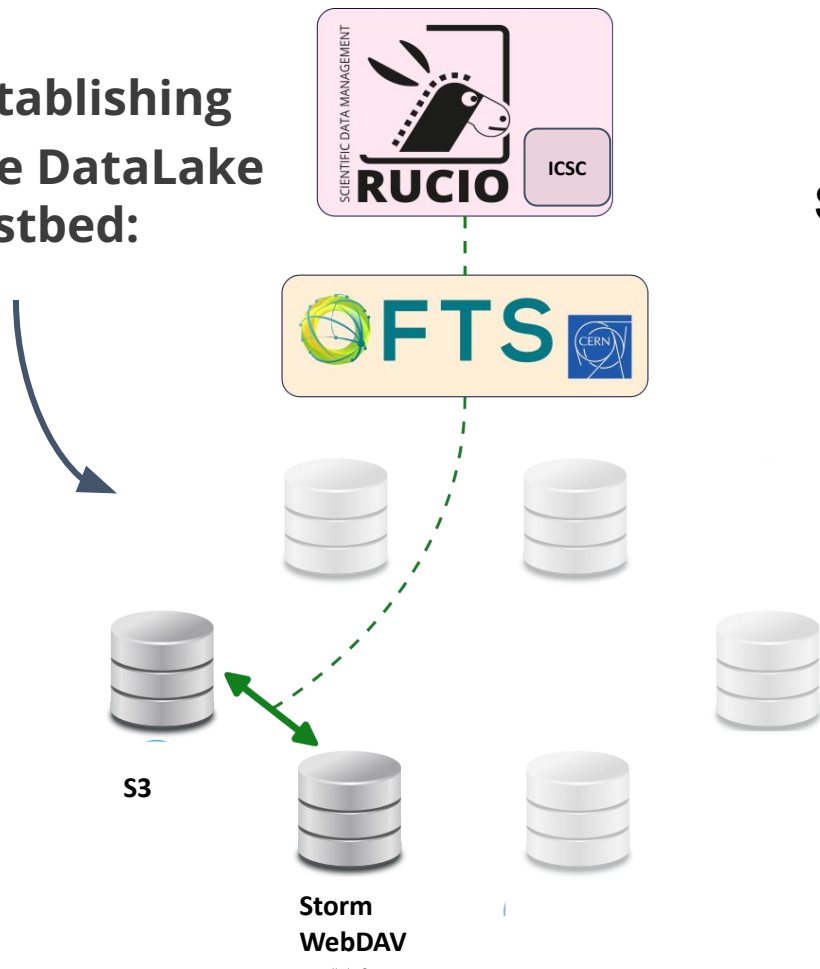


More on [G.Salabella Talk](#)
[A.D'Onofrio Talk](#)

Testbed: what come next (early 2024)

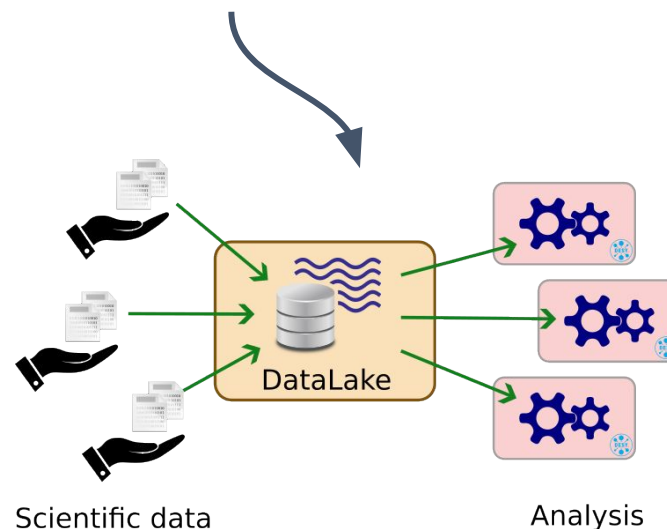
Data Management Services: data orchestration and data transfers

Establishing the DataLake testbed:



Start integration with science use-cases

- Collect exemplar datasets
- The datasets will be injected in the Datalake building science driven playground



Innovation Grant
(Leonardo)

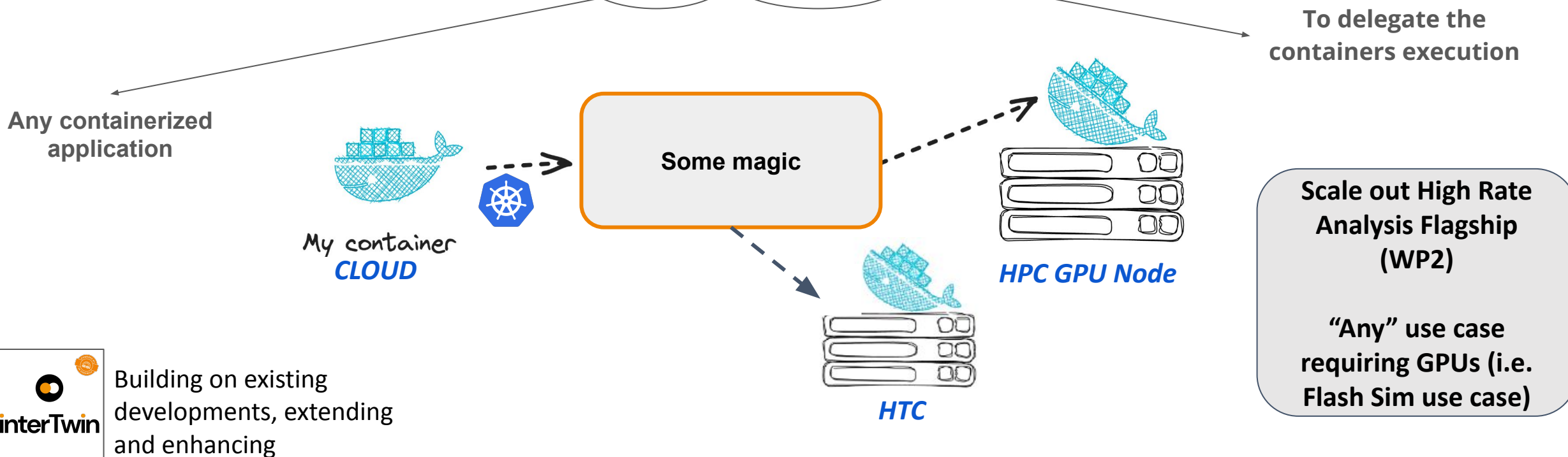
WP3 and WP1 use cases
not yet deeply discussed

Further activities in the pipeline

More on [T.Tedeschi Talk](#)

Transparently extend analysis testbed to run “*any application anywhere*”

- To federate (highly) heterogeneous and disparate ICSC providers
 - enabling a “transparent payload offloading”



Building on existing developments, extending and enhancing

Training Opportunities

We foreseen a new edition of the SOSC (School on Open Science Cloud) during the second half of 2024

- Only in person events and details will be defined in the next 2 months.

We discussed (no follow-up so far) **synergic training with WP4**

- Access to GPUs
- Access and exploitation of FPGA

We will propose a dedicated training on Data Access and Data Sharing (Data Management in the Data Lake)

- This can be in the form of a fully on-line event
 - A webinar?

Interaction with ICSC infrastructure (Spoke 0)

WP5 has several connections with Spoke 0. We foresee our interactions will grow

- “One of the objective of WP5 is to support of the adaptation of existing applications on the data-lake distributed infrastructure”
 - and via innovative computational models

Computing Capacity.

As WP5 we don't expect dedicated resources (indeed no requests to the RAC made).

- We plan to “use” resources from use cases / flagship

All to be implemented

Services Portfolio

We expect to contribute to the evolving portfolio of services:

- Our R&D and co-designed solutions can become generic services
- We expect a process where we can make PR
 - For further evaluations from experts
- We expect to gather feedback (i.e. bi-directional) for all the infrastructural choices we will made/need to take

Summary and lessons

→ The bootstrap phase of the project/WP5 required quite some effort

- Successfully established communication channels and regular meetings
- Defined a work plan for the early phase and contributing to IG and Flagship

→ To improve: The interactions between WPs

- So far interactions come from bottom (not a bad things per se), but it would be very useful to improve the coordination/interactions between WPs on the different common topics
- Not yet worried but **we look forward for Spoke 0 - Spoke 2/WP5 channel**

→ From technical perspective we did the “easy part”. We now enter the “operational phase”

- **This will requires effort** to provide support and accommodate “computing model requirements
- As well as a interactions with infrastructural layer

Backup Slides

WP5: People

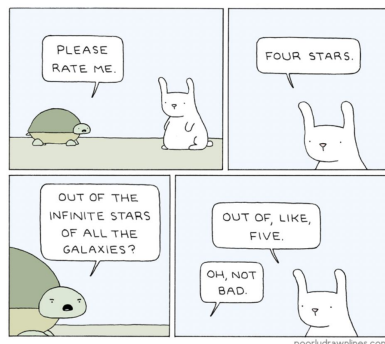
Contributors: INFN, UNIMIB, UNINA, ROMA1, UNITS, UNIBO, UNIPD, UNIFE

Local coordinators

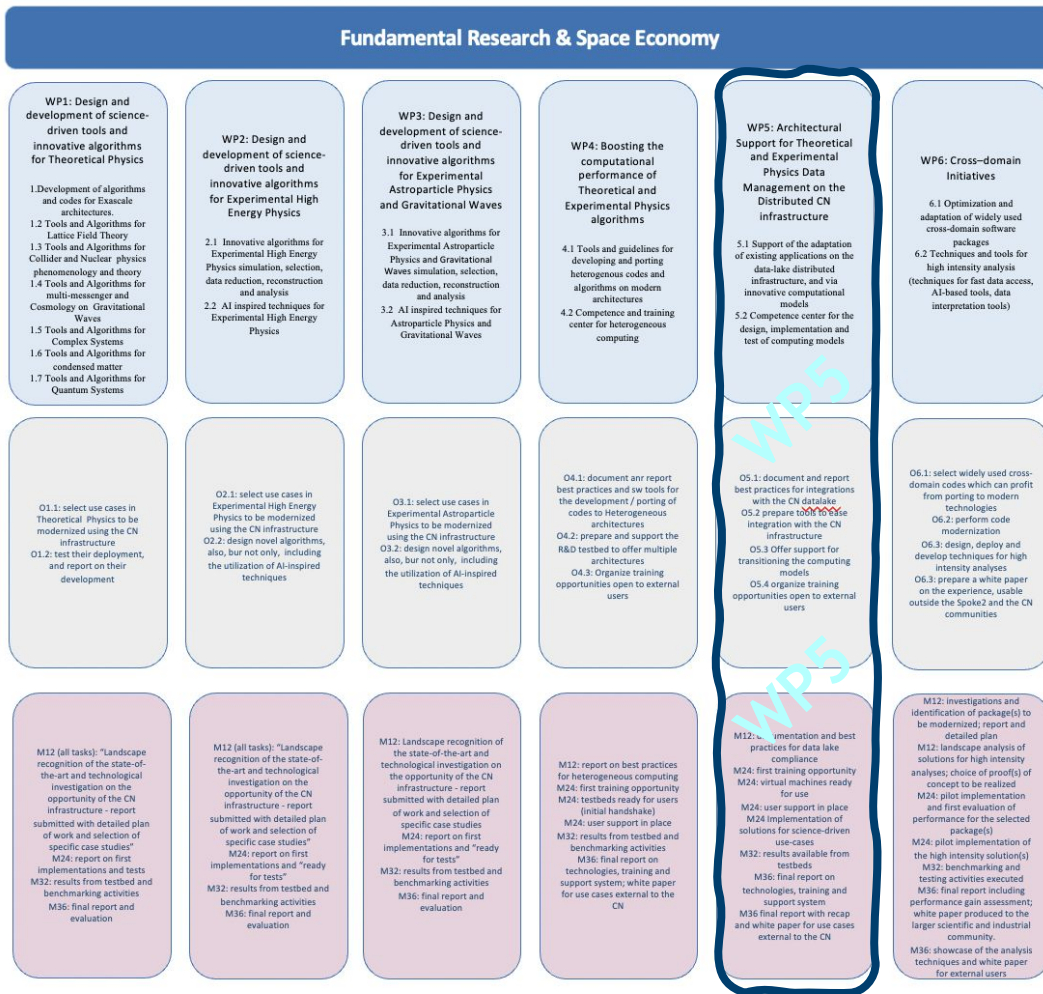
WP5 - People

WP5		
	Nome	Mail
National coordinators	Elvira Rossi Daniele Spiga	elvira.rossi@unina.it daniele.spiga@pg.infn.it
Institute coordinators		
INFN	Daniele Spiga	daniele.spiga@pg.infn.it
UNIMIB	Mattia Bruno	Mattia.Bruno@mib.infn.it
UNINA	Elvira Rossi	elvira.rossi@unina.it
UNITS	Andrea Bressan	andrea.bressan@ts.infn.it
UNIBO	Alessandra Fanfani	Alessandra.fanfani2@unibo.it
UNIFE	Luca Tomassetti	luca.tomassetti@unife.it

Adelina D'Onofrio	Federica Simone	Mattia Bruno
Alessandra Fanfani	Francesco Ciroto	Michele Pavone
Alessandro De Salvo	Francesco Gravili	Muhammad Anwar
Alessia Spolon	Francesco Noferini	Nadia Rega
Andrea Bressan	Francesco Visconti	Orso Iorio
Andrea Contu	Gianluca Sabella	Paolo Dini
Antonio Stamerra	Giovanni Della Ricca	Piergiulio Lenzi
Alessandro Bombini	Giovanni Andronico	Roberto Peron
Bernardino Spisso	Giovanni Ianniruberto	Sandra Malvezzi
Carmelo Magnafico	Giuseppe Milano	Sara Vallero
Constantinos Siettos	Guido Russo (PO)	Simone Gennai
Daniele Bonacorsi	Guido Russo (PU)	Stefano Bagnasco
Daniele Spiga	Lorenzo Rinaldi	Stefano Bagnasco
Domenico Elia	Luca Tomassetti	Tommaso Diotallevi
Eleonora Luppi	Luca Zampieri	Tommaso Tedeschi
Elvira Rossi	Lucia Silvestris	Valentina Fioretti
Fabio Garufi	Marco Landoni	



Spoke 2 - Fundamental Research & Space Economy



Spoke 2 - objectives

The Spoke 2 intends to address the needs of theoretical and experimental physics with accelerators, astroparticle physics with space- and ground-based detectors and gravitational wave investigation designing, developing and testing solutions apt to the current and next-generation experiments, and fitting the opportunities provided by the PNRR and the National Centre (CN) "Big Data, HPC and Quantum Computing".