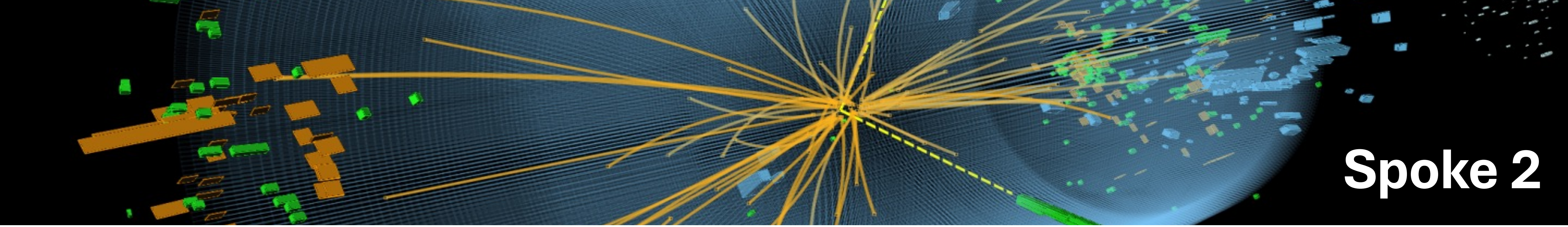
ICSC **Annual Meeting**

del Centro Nazionale HPC, BIG DATA AND QUANTUM COMPUTING

Roma
Auditorium Antonianum

Quasi interactive analysis of big data with high throughput

Tommaso Diotalevi, Università di Bologna
on behalf of Spoke2 members

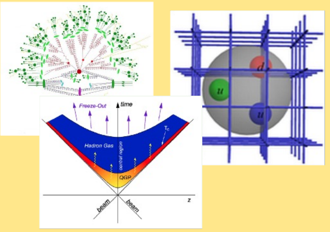


Spoke 2

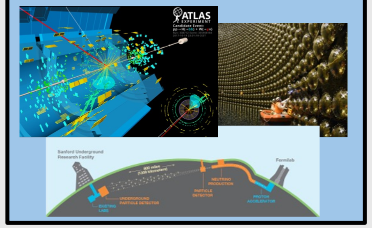
FUNDAMENTAL RESEARCH & SPACE ECONOMY

Scientific

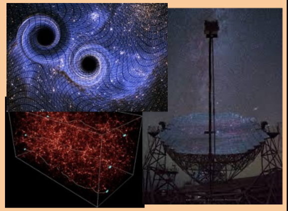
WP1: tools and algorithms for Theoretical Physics



WP2: tools and algorithms for Experimental High Energy Physics



WP3: tools and algorithms for Experimental Astroparticle Physics and Gravitational waves



ICSC-SPOKE2
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Technologic

WP4: tools for porting/optimization on new architectures (low power, GPU, FPGA, ...)

WP5: Boosting computational performance on the distributed CN infrastructure

WP6: cross domain initiatives + space economy

Istitution leader
INFN Istituto Nazionale di Fisica Nucleare

Istitution co-leader
INAF Istituto Nazionale di Astrofisica

Istitutions and Universities

Politecnico di Bari; Sapienza Università di Roma; Università degli Studi di Bari Aldo Moro; Alma Mater Studiorum Università di Bologna; Università della Calabria; Università di Catania; Università degli Studi di Ferrara; Università degli Studi Firenze; Università degli Studi di Milano Bicocca; Università degli Studi di Napoli Federico II; Università degli Studi di Padova; Università del Salento; Università degli Studi di Trieste

Companies

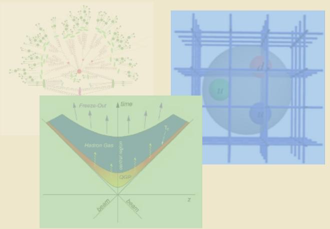
eni; Intesa Sanpaolo; Leonardo; Thales Alenia Space; UnipolSai Assicurazioni; iFAB International Foundation Big Data and Artificial Intelligence for Human Development; sogei

Want to know more about Spoke2 and WP2?

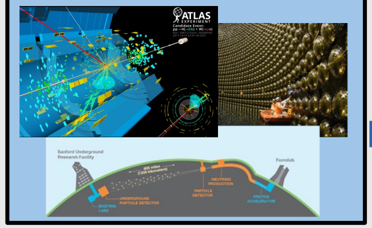
WP2: Talk by **P.Lenzi** and **A.Annovi**, at the Spoke2 meeting of 2023.
Spoke2: Talk by **D.Bonacorsi** (**T.Boccali**, **S.Malvezzi**), at the ISGC 2024 conference.

Scientific

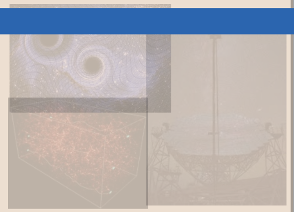
WP1: tools and algorithms for Theoretical Physics



WP2: tools and algorithms for Experimental High Energy Physics



WP3: tools and algorithms for Experimental Astroparticle Physics and Gravitational waves



ICSC-SPOKE2
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

WP4: tools for porting/optimization on new architectures (low power, GPU, FPGA, ...)

alpaka

NVIDIA CUDA

WP6: cross domain initiatives + space economy

GEANT4

copernicus

BLOCK CHAIN

WP5: Boosting computational performance on the distributed CN infrastructure

INFN CLOUD

RUCIO

jupyter

Technologic

supported by

1 of the 19 Spoke2 flagship use cases:

UC2.2.2

Finanziato dall'Unione europea NextGenerationEU | Ministero dell'Università e della Ricerca | Italiadomani | ICSC

Quasi interactive analysis of big data with high throughput

Spoke	2
WP	2, 5
Use case short name	Quasi interactive analysis of big data with high throughput
Use case ID	UC2.2.2
Expected Completion	31/8/2025

Approval workflow

Status	Version	Date	Submitter	Note	Signature
Draft	1.0	03/07/23	WP Leaders	First version	
Final Version	1.1	1/9/2023	WP Leaders		
Approved by Spoke Leaders	1.1	11/9/2023	Spoke Leaders		

Principal Investigators:

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

Tommaso Diotalevi

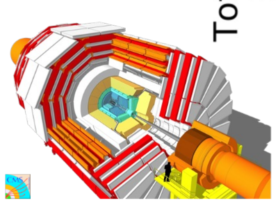
UNIVERSITÀ DEL SALENTO

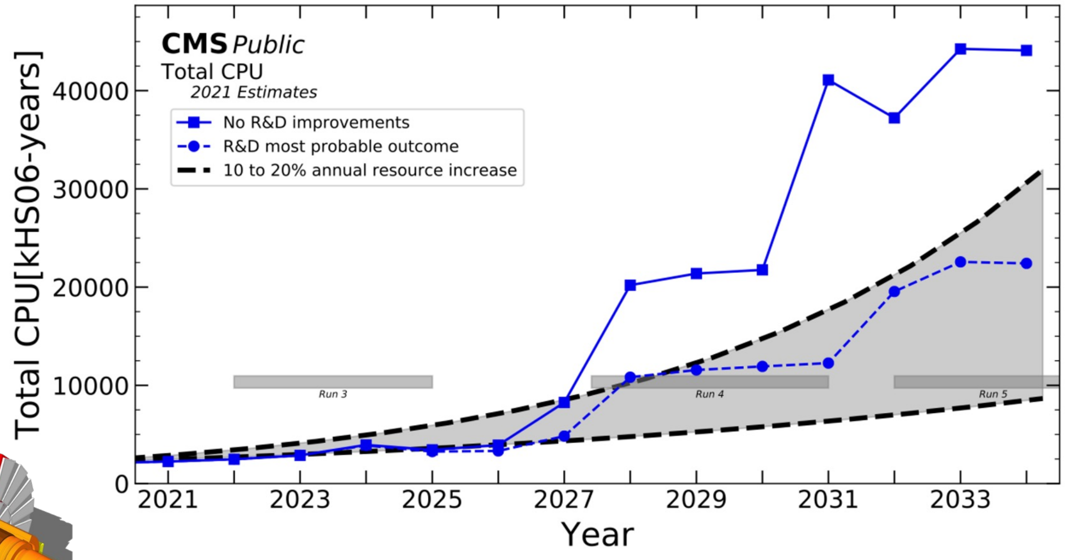
Francesco G. Gravili

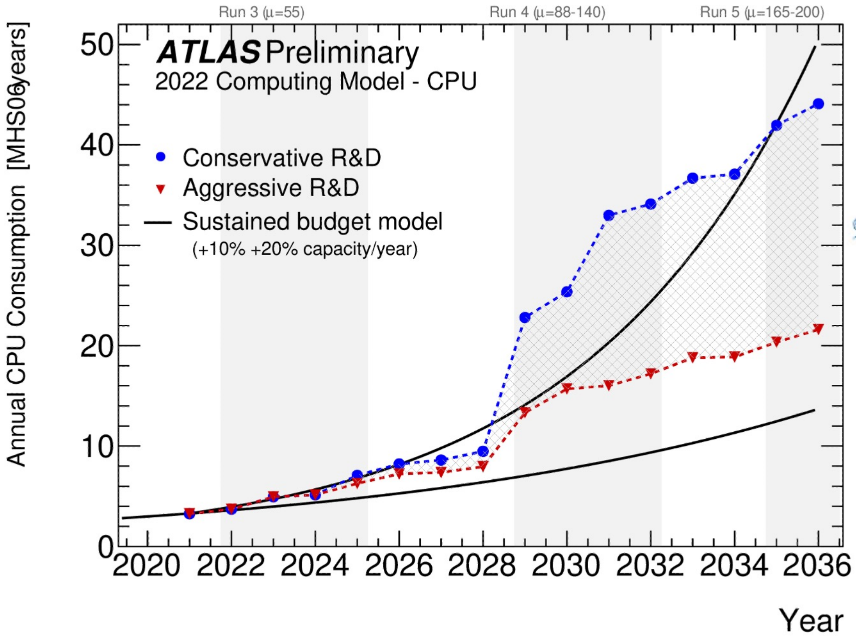
Introduction

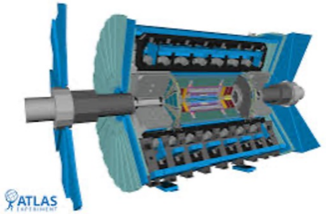
- Analysing large amounts of data efficiently, exploiting the available resources as much as possible, is a common challenge both for research and industry.

- From the beginning, the High Energy Physics (HEP) experiments at CERN, gave much attention to the computing and data management aspects. Nevertheless, the **next phases of the Large Hadron Collider** (HL-LHC) will require an even greater effort.









Introduction

Some estimate for the next 5-10 years of CMS operation:

- ~30 Billion collision events + 30 Billion simulation events;
- Each event: 2-4 kB;
- The last update of the CMS Computing model foresees this throughput:

Name	Length	% of the dataset	Data to process	Event, data rate
“A coffee”	< 5 min	1% (~0.6B evts)	~2 TB	~1.7MHz, ~7GB/s
“A lunch break”	1 hour	10% (~6B evts)	~20 TB	~1.5MHz, ~6GB/s
“A night”	12 hours	100% (60B evts)	~200 TB	~1.2MHz, ~5GB/s

- Difficult to get more than 100 Hz/CPU core → needs efficient distribution on a few tens of machines;

Not only concerning the HEP domain (“Data is data”):

- More and more scientific / industrial / societal domains have or will have soon needs similar to those from LHC:

DEEP UNDERGROUND NEUTRINO EXPERIMENT

ProtoDune: 2-3GB/s (like CMS); Real Dune: 80x

SKA: up to 2 PB/day;

A single genome: ~100GB, a 1M survey=100PB

CTA projects: up to 10PB/y

New analysis paradigm based on:

- Declarative programming and interactive workflows;
- Distributed computing on geographically separated resources.

Introduction

Some estimate for the next 5-10 years of CMS operation:

- ~30 Billion collision events + 30 Billion simulation events;
- Each event: 2-4 kB;
- The last update of the CMS Computing model foresees this throughput:

Name	Length	% of the dataset	Data to process	Event, data rate
"A coffee"	< 5 min	1% (~0.6B evts)	~2 TB	~1.7MHz, ~7GB/s
"A lunch break"	1 hour	10% (~6B evts)	~20 TB	~1.5MHz, ~6GB/s
"A night"	12 hours	100% (60B evts)	~200 TB	~1.2MHz, ~5GB/s

- Difficult to get more than 100 Hz/CPU core → needs efficient distribution on a few tens of machines;

Not only concerning the HEP domain ("Data is data"):

- More and more scientific / industrial / societal domains have or will have soon needs similar to those from LHC:

ProtoDune: 2-3GB/s (like CMS); Real Dune: 80x

SKA: up to 2 PB/day;

A single genome: ~100GB, a 1M survey=100PB

CTA projects: up to 10PB/y

High Throughput Platform

New analysis paradigm based on:

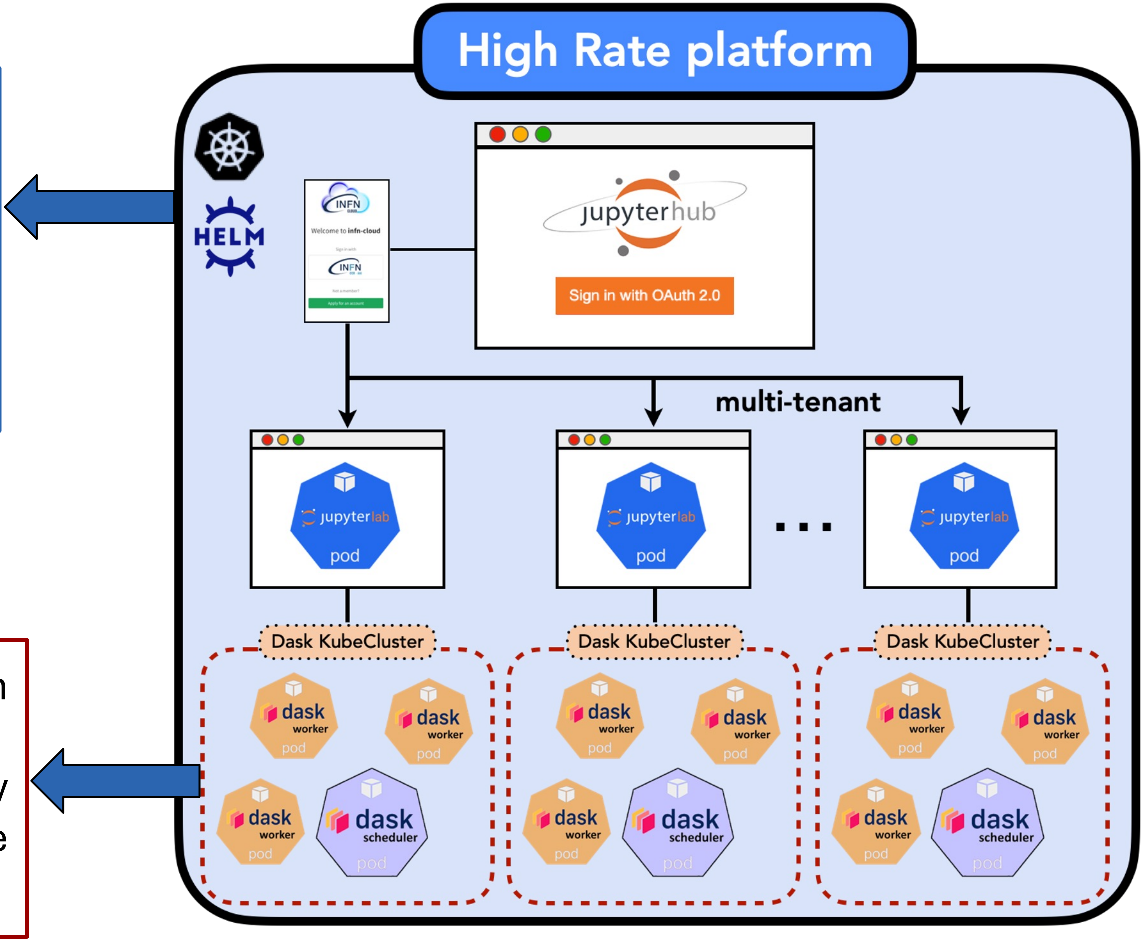
- Declarative programming and interactive workflows;
- Distributed computing on geographically separated resources.

High Throughput platform

Deployment of the Kubernetes resources handled via HELM Charts.

↓

Scalable on available resources

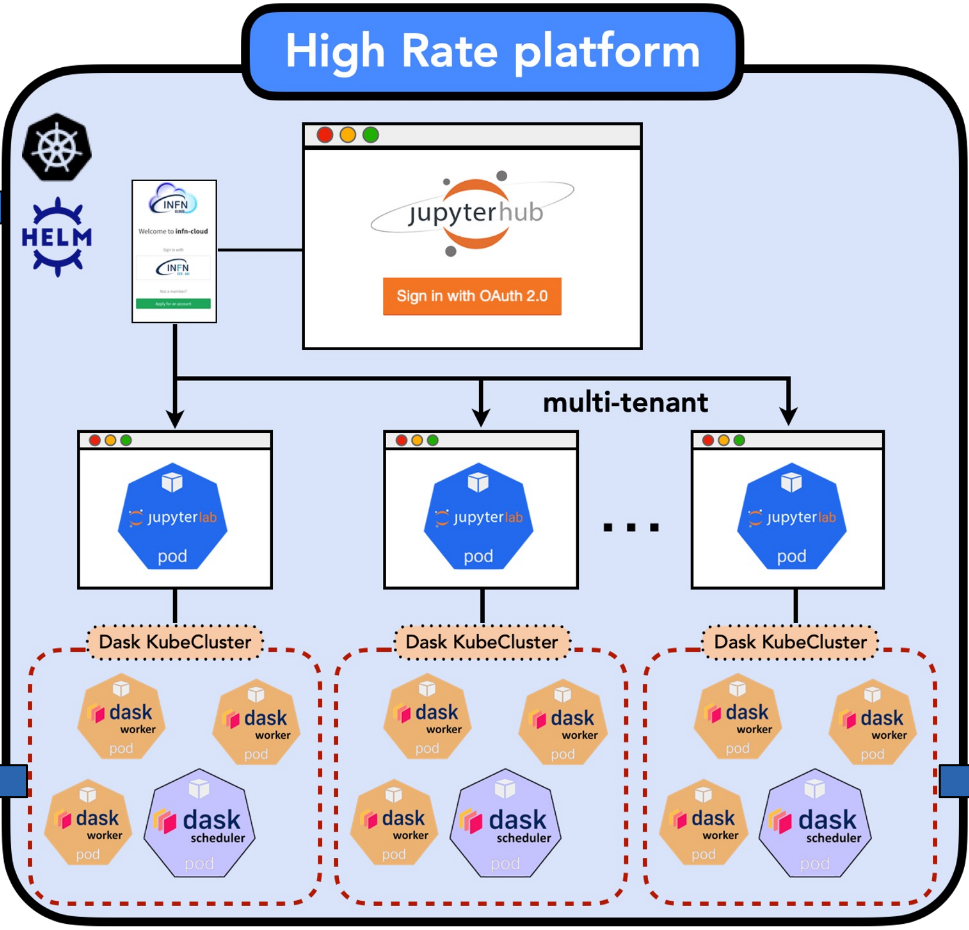


- The execution happens in the Dask Cluster;
- Users choose on how many cores parallelly distribute the analysis.

High Throughput platform in ICSC

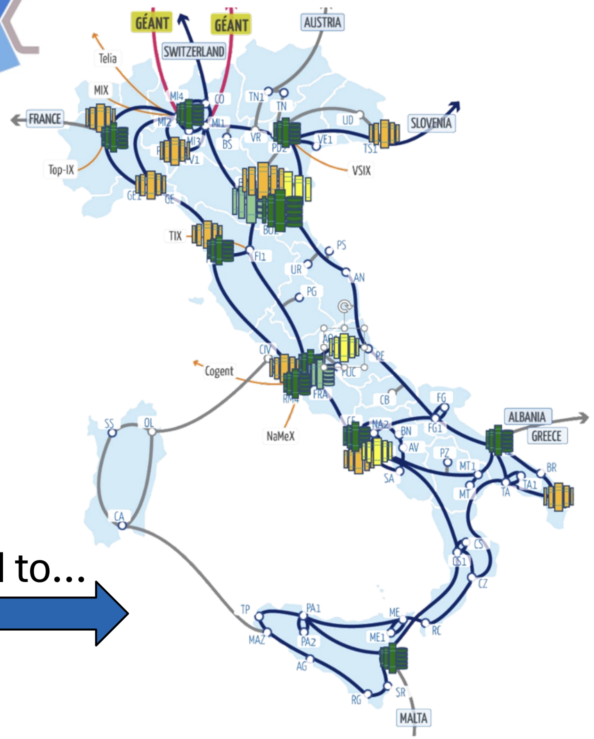
Deployment of the Kubernetes resources handled via HELM Charts.

Scalable on available resources



- The execution happens in the Dask Cluster;
- Users choose on how many cores parallelly distribute the analysis.

offloaded to...



Re-thinking the analysis pipeline

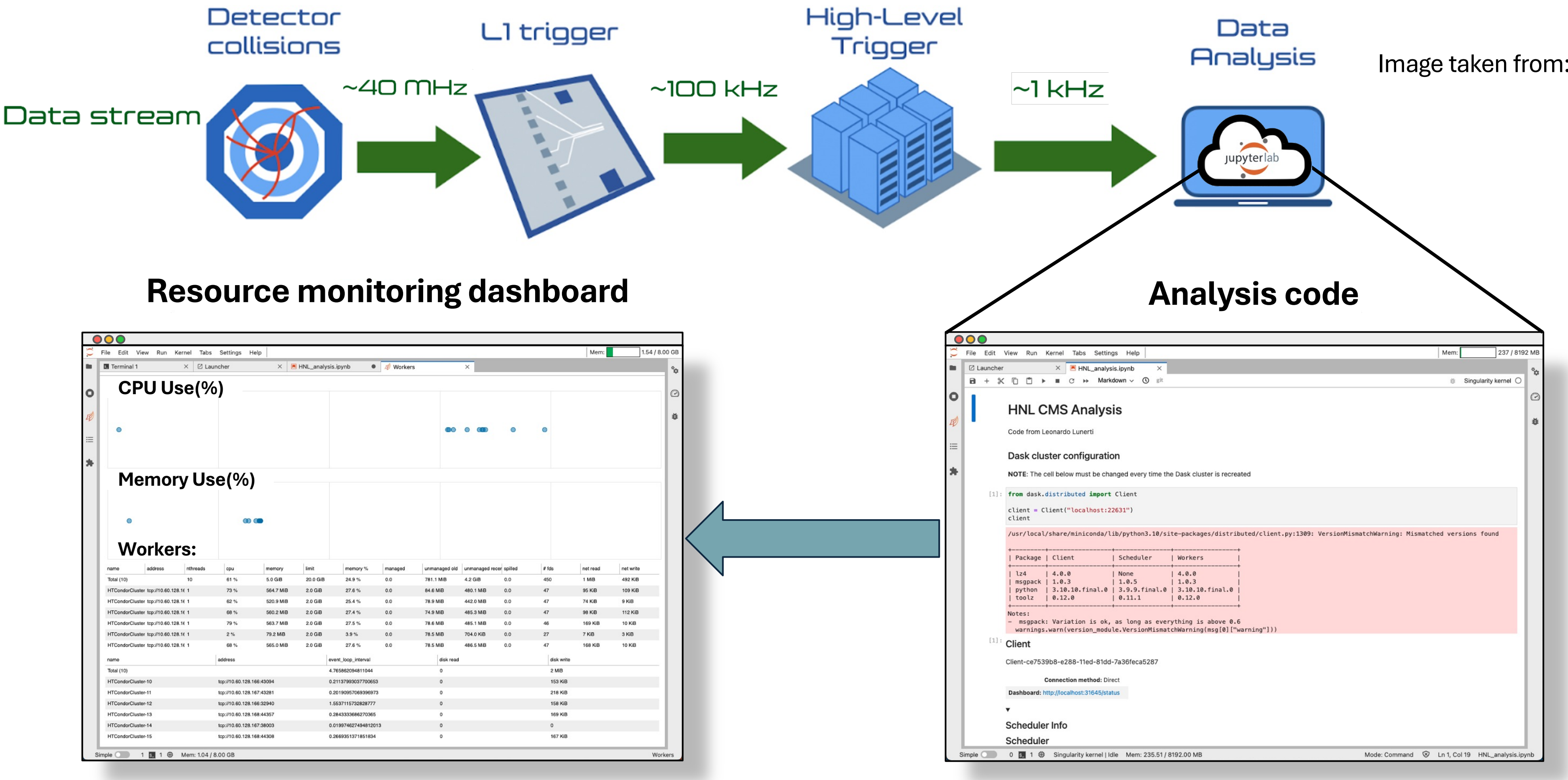
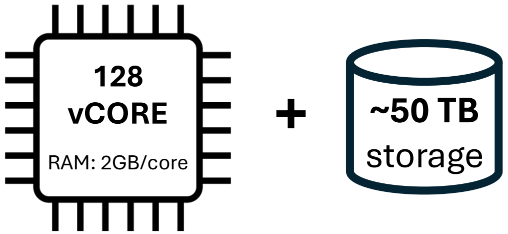


Image taken from: [link]

ICSC resources required by the flagship

Resources submitted to RAC on spring, **and recently provisioned**.

- **Current phase** (deployment of the cloud infrastructure):



(For the dynamic allocation of volumes, for user working areas.)

The first analyses porting, using a prototypal platform running on these resources, is undergoing.

- **Next phases**: up to 670 cores CPU for the analyses scale tests, moving towards the finalization of the infrastructure (by the end of the project).

* The adoption of *heterogeneous resources (i.e. GPUs), in the near future, is not excluded, based on possible synergic applications with other flagship activities.*

Activities (so far) orbiting around the flagship

Vector Boson Scattering ssWW analysis in hadronic tau and light lepton

Heavy Neutral Lepton search on heavy neutrinos in the D_S decays

Muon detector performance analysis

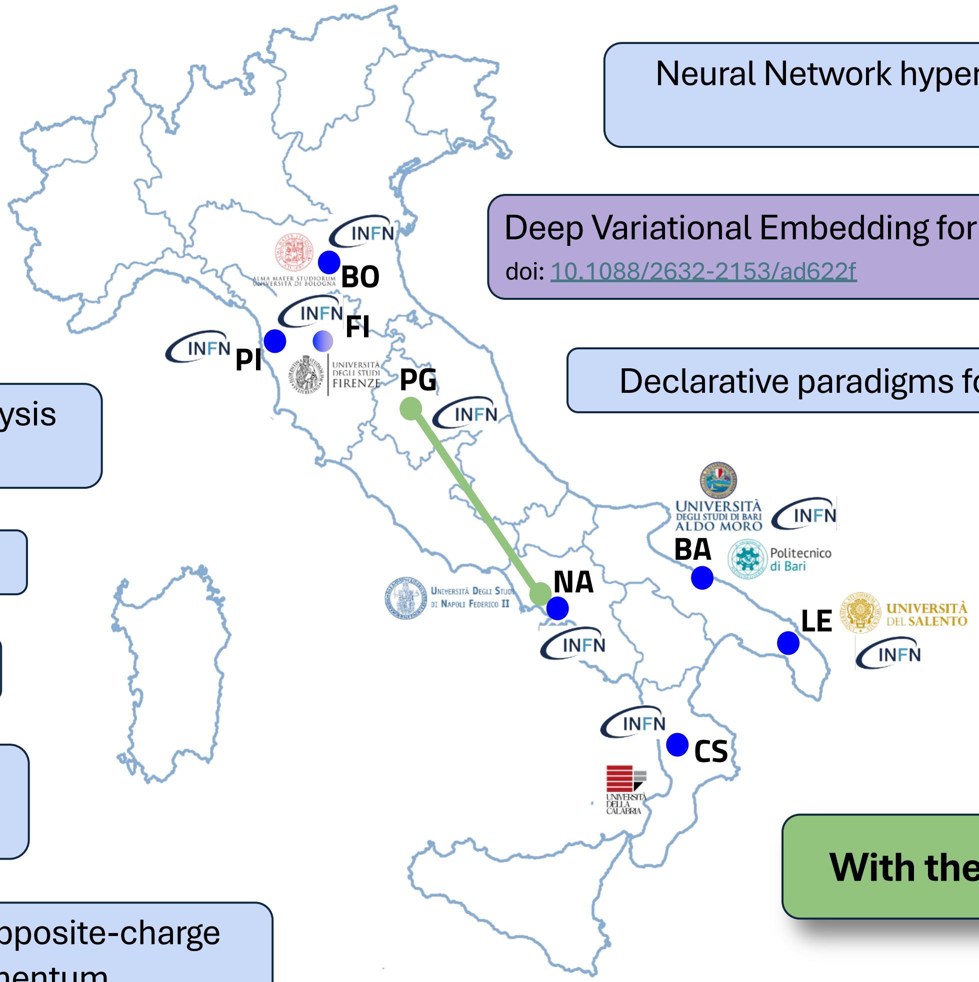
Continuous Integration pipeline, triggering analysis execution on AF

di-Higgs decaying to two b quarks and two muons

Search of rare events in tau to 3 muons decay

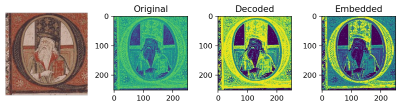
Differential cross section measurement for ttbar inclusive production

Search for new phenomena in events with two opposite-charge leptons, jets and missing transverse momentum



Neural Network hyperparameter optimisation applied to future colliders (FCC-ee)

Deep Variational Embedding for Cultural Heritage
doi: 10.1088/2632-2153/ad622f



Declarative paradigms for analysis description and implementation

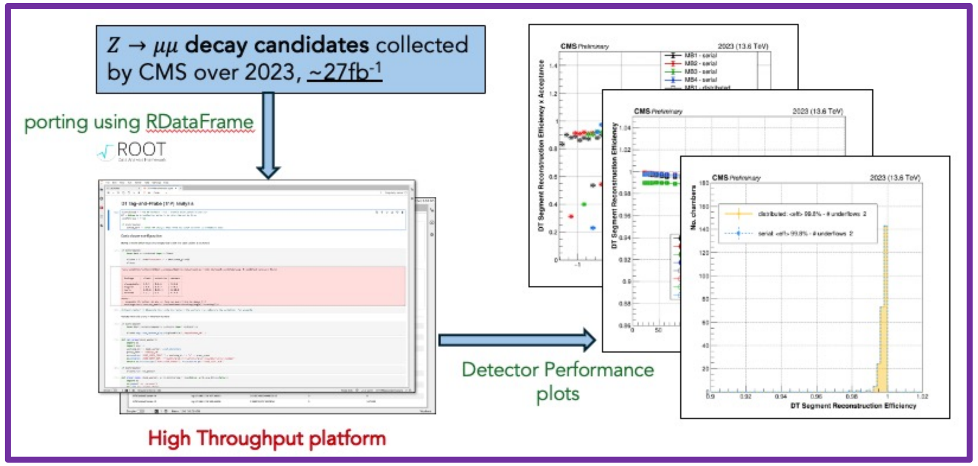
top quark+MET analysis

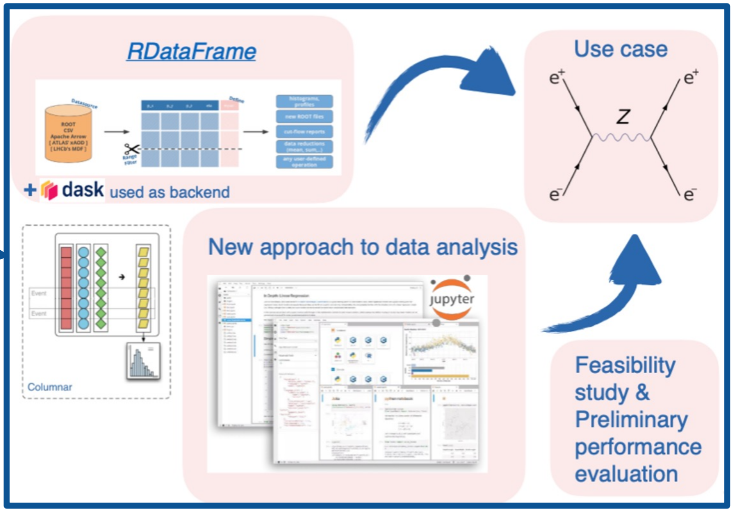
Benchmark interactive analysis for future colliders (FCC-ee)

With the infrastructural support of WP5

Scientific production in conferences

- Poster at the "International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT 2024)":
  - *Declarative paradigms for analysis description and implementation.*
  - *Quasi interactive analysis of High Energy Physics big data with high throughput.*
- Talk at the "Incontri di Fisica delle Alte Energie (IFAE 2024)":
  Analisi quasi-interattiva per big data con alto throughput per la Fisica delle Alte Energie.
- Talk at the "International Conference on High Energy Physics (ICHEP 2024)":
  Enhancing CMS data analyses using a distributed high throughput platform.
- Talk at the "2nd European Committee for Future Accelerator (ECFA) Workshop on Higgs/EW/Top Factories":
  Benchmark interactive analysis for future colliders.
- Talk at the "Conference on Computing in High Energy and Nuclear Physics (CHEP 2024)":
  Leveraging distributed resources through high throughput analysis platforms for enhancing HEP data analyses.





ICSC 2024 Annual Meeting

Conclusions

- The challenge presented by the next LHC phases requires a strong development effort of new tools, for making data analysis as efficient and as modern as possible;
- In synergy with the big collaborations at CERN, a new **High Throughput Platform** has been developed:
  - Based on *interactive workflows* and on *declarative programming*;
  - Running on distributed resources (and heterogeneous).
- Several analysis from the HEP world are already testing such infrastructure, for performance measurements;
- Thanks to the resources allocated by RAC, the first tests with ICSC resources are undergoing.
- Strong synergy with other Work Packages inside Spoke2: both in "*scientific*" and "*technological*" aspects.

Once fully operational, such platform will be used by the **entire ICSC community**, including all kinds of industrial applications.

Thank you for the attention!

KPI– Key Performance Indicators

KPI ID	Description	Acceptance threshold
KPI2.2.2.1	Implementation of N data analyses in the AF	$N \geq 2$
KPI2.2.2.2	Reference documentation of the AF	≥ 1 dedicated web site
KPI2.2.2.3	Hands-on workshops for AF users	≥ 1 workshops
KPI2.2.2.4	Scaling up the testbed AF infrastructure, serving k tenants, for a total of N data analyses	$\geq (200 \cdot M)$ cores
KPI2.2.2.5	Talks at conferences/workshops about AF activities	≥ 1 talk

RAC resources detail

7.1 Resources granted by INFN-CLOUD (PaaS)

vCPU (number of vCores and requested allocation time)

Number *	Time *	vCORE *	Notes
4	17520	32	time in hours

Number of requested GPU and allocation time

Number *	Time *	GPU *	Notes
1	1	1	NO GPU REQUIRED. The numbers were entered due to a problem with the form

vCPU (number of vCores and requested allocation time)

RAM per VCORE *	Notes
4	Memory in GB

- Software used or required, including preferred Cloud services

I.e.: Kubernetes-as-a-service, Jupyter Notebook as a Service, Private Container Image Registry, Spark and Grafana as a service, Dropbox-like sync-and-share service

Software used or required	Notes
	Total core hours: 4*17520*32 = 2242560. This kind of resource does not follow time evolution, to be considered constant for the entire project duration.

7.2 Resources granted by INFN-GRID (Batch processing)

CPU (number of Cores and requested allocation time)

Number *	Time *	vCORE *	Notes
21	17520	32	time in hours

RAM Requirements

RAM per VCORE *	Notes
4	Memory in GB

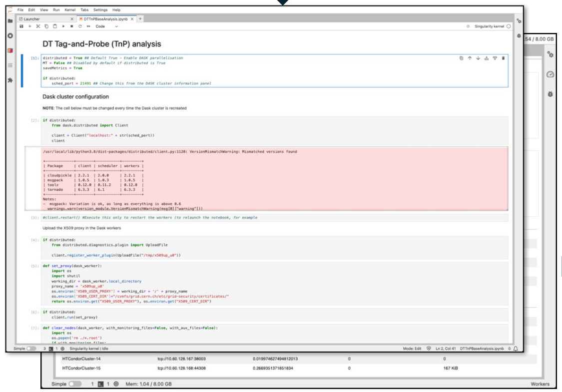
Software used or required	Notes
	Total core hours: 21*17520*32 = 11773440. This kind of resource follows a time evolution, where: - first 6 months: 876000 core hours, to be considered as 1000 cores * 20% of time (1000 cores distributed in 5 data analyses); - following 4 months: 730000 core hours, to be considered as 1000 cores * 25% of time (1000 cores distributed in 5 data analyses) - following 4 months: 730000 core hours, to be considered as 1000 cores * 25% of time (1000 cores distributed in 5 data analyses) - last 10 months: 9437440 core hours, to be considered as 5000 cores * ~26% of time (5000 cores distributed in 5 data analyses) These nodes, on N CPU systems without the requirement of fast inter-node infiniband communication (HTC-like), must have access to permanent storage.

Lorem ipsum dolor sit amet, consectetur adipiscing elit,

$Z \rightarrow \mu\mu$ **decay candidates** collected by CMS over 2023, ~27fb^{-1}

porting using RDataFrame

ROOT Data Analysis Framework



High Throughput platform

Detector Performance plots