Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

WP2: Design and development of science-driven tools and innovative algorithms for Experimental High Energy Physics

Piergiulio Lenzi
Alberto Annovi

ICSC and Spoke 2, 11 dicembre 2024

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Scientific organization

5 main research lines, called flagship use cases, listed below with their PIs.

- UC2.2.1 Advanced ML: flash simulation and other bleeding edge applications
  - ✓ Lucio Anderlini (INFN Fi)
- UC2.2.2 Quasi interactive analysis of big data with high throughput
  - ✓ Tommaso Diotalevi (UniBo), Francesco G. Gravili (UniSalento)
- UC2.2.3 Development of ultra-fast algorithms running of FPGAs
  - ✓ Bernardino Spisso (UniNa), Simone Gennai (INFN MiB)
- UC2.2.4 Porting of algorithms to GPUs
  - ✓ Adriano Di Florio (Poliba)
- UC2.2.5 Physics validation of reconstruction code on ARM
  - ✓ Francesco Noferini (INFN Bo)

Many thanks to the PIs and all the WP2 members for their continued effort in bringing these topics forward.

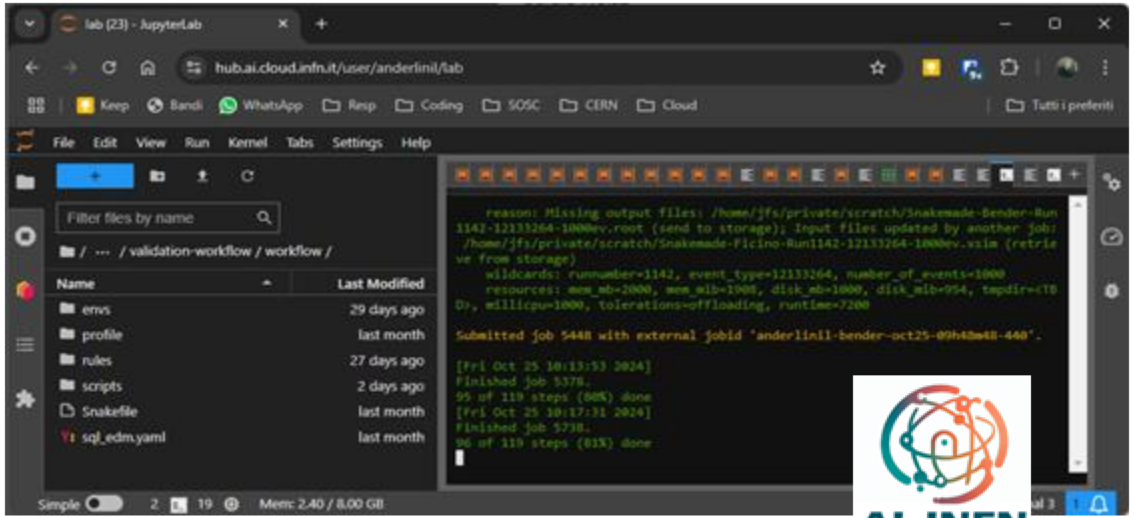
Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

UC2.2.1 Flagship activity: Advanced Machine Learning. Flash Simulation and bleeding edge applications

Develop and showcase the distributed computing infrastructure for advanced applications of Machine Learning.

Breadth*: many use cases can benefit from the infrastructure*

Depth*: LHCb Flash Simulation training and validation can be entirely deployed on the new infrastructure.*



AI_INFN

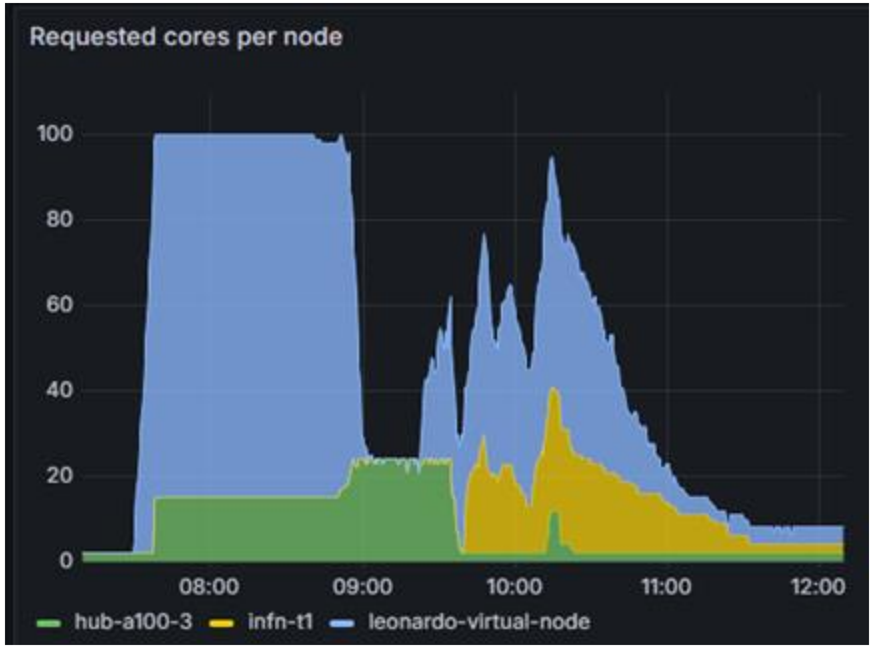
HTC

interLink

HPC payloads

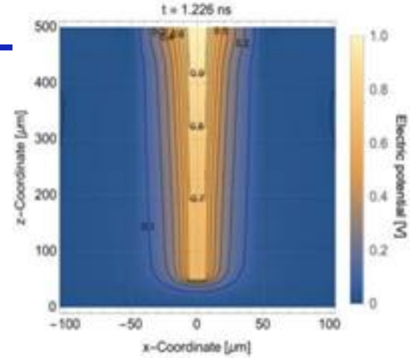
UC2.2.1 Flagship activity: Advanced Machine Learning. Flash Simulation and bleeding edge applications

Flash Simulation Validation pipeline running for the first time on the distributed infrastructure as part of the ***"Integration PoC"***

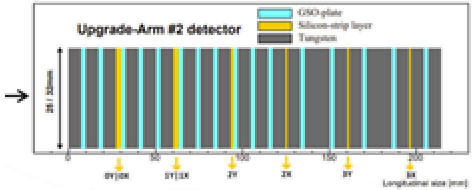


Other applications part of the Flagship

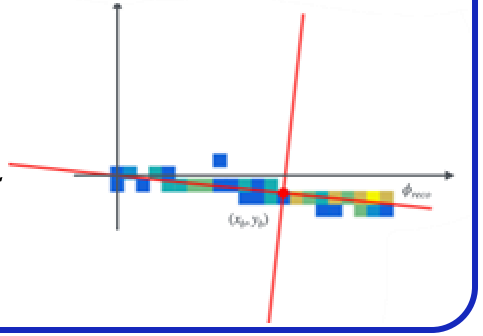
- *Simulation of 3D diamond detectors with pseudospectral and PIML techniques (see **Alessandro Bombini**'s talk, Firenze)*



- *Reconstruction of calorimetric clusters using Machine Learning for the LHCf experiment at CERN (**Giuseppe Piparo**, Catania)*



- *Proton-recoil track imaging detector for fast neutrons (**Samuele Lanzi**, CNAF)*



UC2.2.1 Advanced ML: flash simulation and other bleeding edge applications

- **KPI - Key Performance Indicator**

KPI ID	Description	Acceptance threshold	Completed %
KPI2.2.1.1	N_{MC} billion events obtained from ML-based simulation, as demonstrated by official links in experiment simulation databases	$N_{MC} \geq 1$	0.1 %
KPI2.2.1.2	N_{EXP} experiments have tested a machine-learning based simulation	$N_{EXP} \geq 2$	150 %
KPI2.2.1.3	Machine-learning use-cases tested in the context of the CN were presented at N_{CONF} international and national events	$N_{CONF} \geq 3$	567%
KPI2.2.1.4	N_{UC} different machine-learning use-cases were tested in the context of the CN and made available in git repositories	$N_{UC} \geq 5$	100 %

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

UC2.2.2 Quasi interactive analysis of big data with high throughput

- UC 2.2.2 document available on GoogleDoc (including KPI table)
- Official mailing list: cn1-spoke2-wp2-analysisfacility@lists.infn.it
- **Several analysis already implemented or in ongoing state (list not fully comprehensive):**
  - **ATLAS**: SUSY search in events with two opposite-charge leptons, jets and missing transverse momentum, using LHC Run2 data
  - **CMS**: Muon detector performance analysis, Search for LFV decays $\tau \rightarrow 3\mu$
  - **FCC-ee**: Reconstruction and scalability tests at Z-pole
  - **Others**: Declarative paradigms for analysis description and implementation, Continuous Integration pipelines
- Still to define assessment criteria to evaluate improvements
- **Several contributions at major Italian and International Conferences: CHEP, ICHEP, IFAE, SIF**

UC2.2.2 Quasi interactive analysis of big data with high throughput

- **The ICSC facility is available**:
  - Technical configuration available on the official ICSC-Spoke2 repository (**maintained by WP5**).
  - **Entrypoint**: https://hub.131.154.98.51.myip.cloud.infn.it/
  - Registrations to the new **IAM-ICSC** service ongoing, which will replace current **IAM-DEMO** service
- The official documentation is available at the official Spoke2 GitHub repo
- Some aspects still to be determined, e.g. data management, *cvmfs* and other ones. First Mini-Workshop in July
- **First batch of resources now available from the RAC:**
  - 128 vCORE (with 2GB RAM per core) and 50TB of storage (for user areas)
  - Plan is to extend such pool, up to 670 vCORE

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

UC2.2.2 Quasi interactive analysis of big data with high throughput

ICSC Workshop on Analysis Facilities

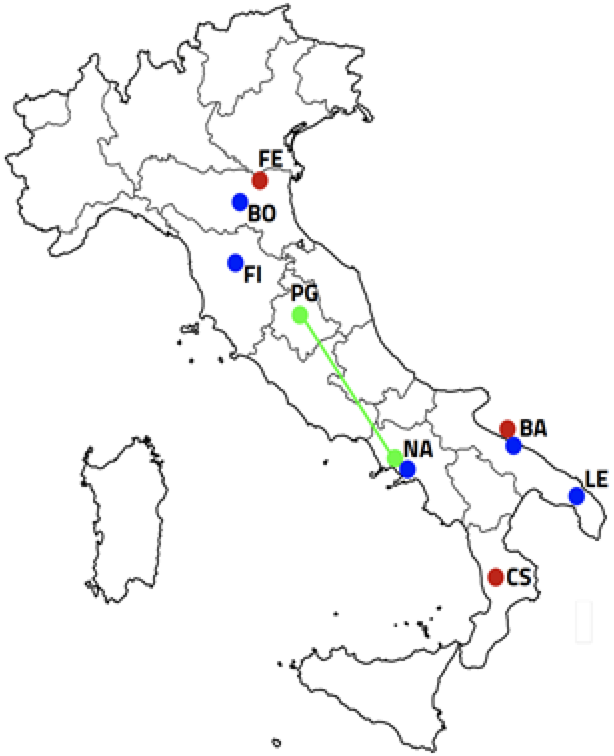
- Location: Bologna (in presence and online)
- Agenda: https://agenda.infn.it/event/44199
- Dates: 8-10 January 2025
- Format:
  - First part open to **everyone:** lectures and hands-on on distributed data analysis
  - Second part **restricted** to experiment communities: specific analyses and future perspective

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

UC2.2.2 Quasi interactive analysis of big data with high throughput

- **KPI - Key Performance Indicator**

KPI ID	Description	Acceptance threshold	Status
KPI2.2.2.1	Implementation of N data analyses in the AF	$N \geq 2$	>100%
KPI2.2.2.2	Reference documentation of the AF	≥ 1 dedicated web site	100%
KPI2.2.2.3	Hands-on workshops for AF users	≥ 1 workshops	0% (Jan 2025)
KPI2.2.2.4	Scaling up the testbed AF infrastructure, serving k tenants, for a total of N data analyses	≥ (200·N) cores	32% (1 of 2 batches)
KPI2.2.2.5	Talks at conferences/workshops about AF activities	≥ 1 talk	>100%



UC2.2.3 Development of ultra-fast algorithms running of FPGAs

- **Use cases, not all of them participate to the KPIs:**
- Trigger, DAQ and on-line processing
  - Development of algorithms based on neural networks and implementation on FPGAs, with application for trigger and anomaly detection at event level and object level for the Atlas experiment.
  - Development of a track reconstruction algorithm, at 30 MHz, on FPGA for LHC-b data acquisition.
  - Development of digital trigger logic for a "missing energy" experiment with a positron beam at CERN (POKER/NA64)
    - Superseded by the OpenCall: AI-supported real-time data reduction algorithms for streaming readout systems.
  - Development of quantum-inspired Tree Tensor Networks for classification in Trigger on FPGA
  - Di-tau trigger development for the CMS Level-1 trigger system
  - Scouting and processing of Level-1 trigger data using FPGA to run on-the-fly momentum object calibration with ML based algorithms
- Developing FPGA tools
  - Development of a Customizable Framework for Multi-FPGA Accelerator Generation via architectures
  - Development and testing of RDMA over converged ethernet (ROCE) on FPGA for data transfer from detectors' front-end to computing servers

Simone.Gennai@cern.ch

Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

UC2.2.3 Development of ultra-fast algorithms running of FPGAs

- **KPI - Key Performance Indicator**

KPI ID	Description	Acceptance threshold	Status
KPI2.2.3.1	Development of triggering algorithms, on-line analyses, data acquisition on FPGA	Submission of 1 paper to a peer-reviewed journal	1 paper already accepted
KPI2.2.3.2	Online scouting	Submission of 1 paper to a peer-reviewed journal	Abstract being submitted to ichep 2024 about scouting
KPI2.2.3.3	Development of tools to integrate several FPGAs together	Submission of 1 paper to a peer-reviewed journal	G. Bortolato et al 2024 JINST 19 C03038
KPI2.2.3.4	Organizing courses about FPGA programming on low and high level	At least two courses organized	1 course done at the end of 2023 1 VHDL course done in February 2024

Simone.Gennai@cern.ch

UC2.2.4
Porting of algorithms to GPUs

	Description	(thr)
KPI2.2.4.1	At least XX offline algorithm ported to GPU (most probably an LHC algorithm)	2 (1)
KPI2.2.4.2	At least YY online algorithm ported to GPU (most probably an LHC algorithm)	2 (1)
KPI2.2.4.3	Preparation of a test infrastructure able to test codes on heterogeneous systems. At least ZZ architectures to be supported (eventually, AMD, nVIDIA, CPU)	3 (3)
KPI2.2.4.4	Organize at least KK events to introduce students and collaborators to heterogeneous computing and train them to the usage of portability tools (joint with WP4).	2 (3)

Flagship targets:

- Heterogeneous computing training (joint with WP4).
- Extending the usage of GPU/accelerators
- Studying the different portability solutions available.

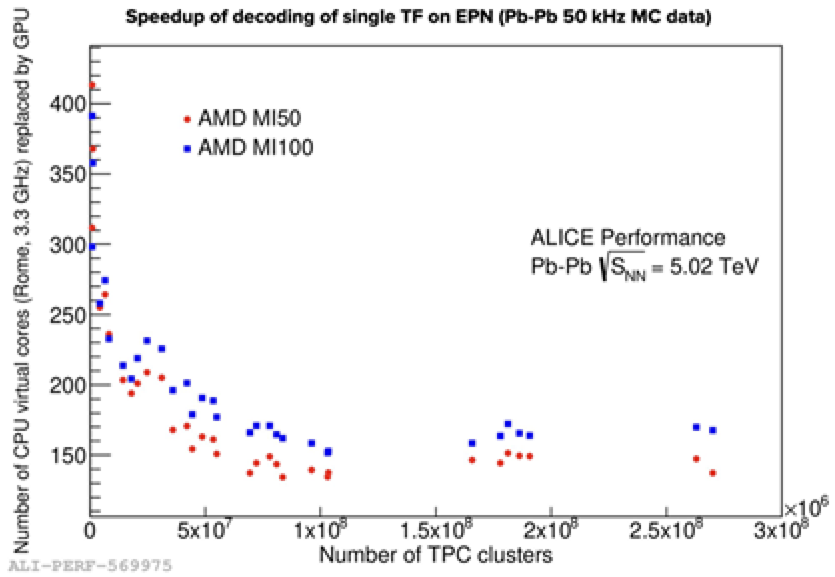
2024 timeline (MS from 7 to 9):

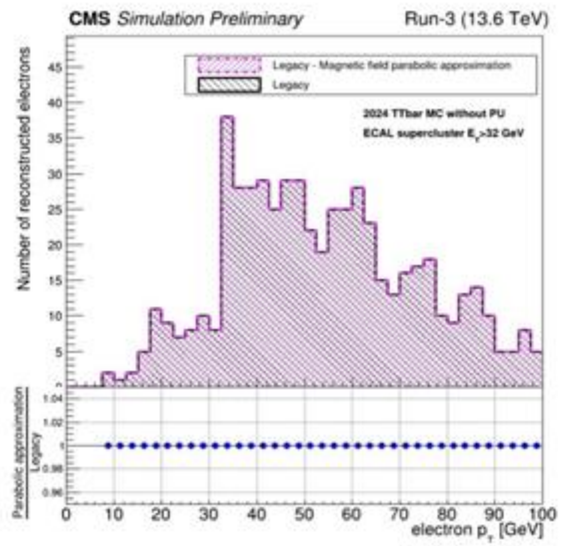
- ❑ Assessment of the different activities being carried within the involved experiments and the flagship participant institutions
- ❑ Algorithms' porting activities leveraging on the experience acquired and on the best practices defined by WP4.
- ❑ Setup of the physics validation infrastructure and definition of the minimal targets, in terms of physics performances and reproducibility

UC2.2.4 Porting of algorithms to GPUs

Some of the activities contributing to KPIs:

- **ALICE TPC data decompression** -> KPI2.2.4.2;
  - The TPC decoding algorithm has been successfully ported to run on AMD and NVIDIA GPUs.
- **CMS Electron Seeding on GPU** -> KPI2.2.4.1 + KPI2.2.4.2;
  - a new parallel algorithm for seeding the electron reconstruction has been developed and will be tested to go in production in 2025
- **CMS Primary Vertex Reconstruction on GPU** -> KPI2.2.4.1;
  - a new GPU-compliant CPU implementation developed for Vertex reconstruction showing a factor five of gain in timing. A first GPU implementation of the algorithm directly in Alpaka portability framework being written.
- **CMS Pixel and Strip Reconstruction on GPU** -> KPI2.2.4.1 + KPI2.2.4.2
  - Pixel track reconstruction has been ported to Alpaka and is in production since 2024.
  - The extension to the strip detector under testing for production.

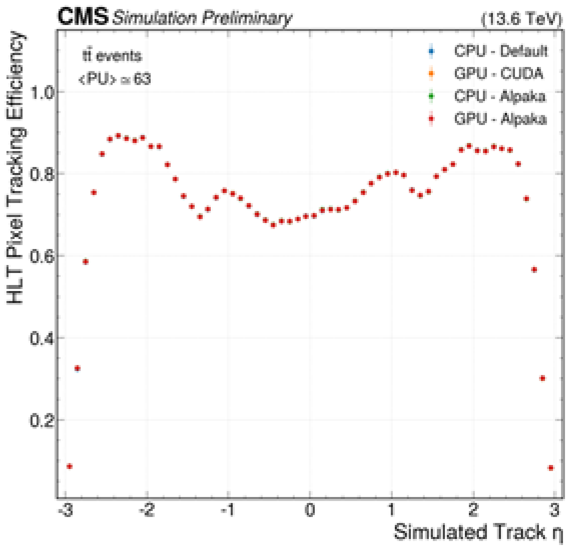


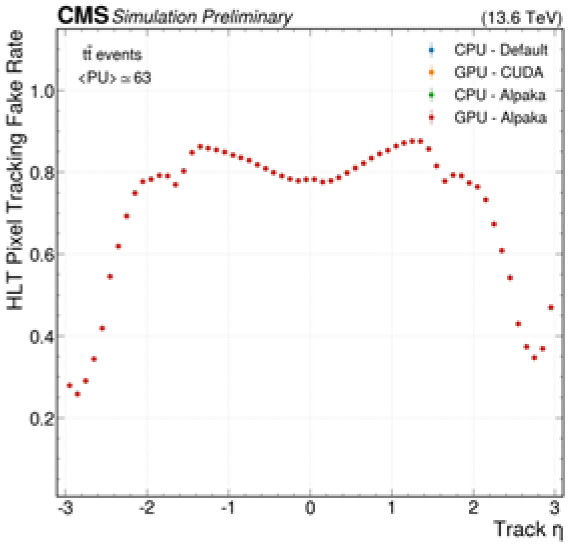


UC2.2.4 Porting of algorithms to GPUs

Validation infrastructure

- working validation infrastructure, tested also on pledged resources;
- validate the usage of Alpaka by CMS from 2024 datataking;
- pixel tracks performance as metrics;
- two types of validations:
  - **Validation 1:**
    - Running on the same RAW events.
    - Four setups (pre-Alpaka and Alpaka)
    - perfect match
  - **Validation 2:**
    - running the Alpaka (and CUDA) GPU and CPU reconstruction in the same job and for each event;
    - compare GPU and CPU quantities event by event;
    - same intrinsic fluctuations in Alpaka and native CUDA.
- **Testing on AMD resources at CNAF for an additional architecture.**



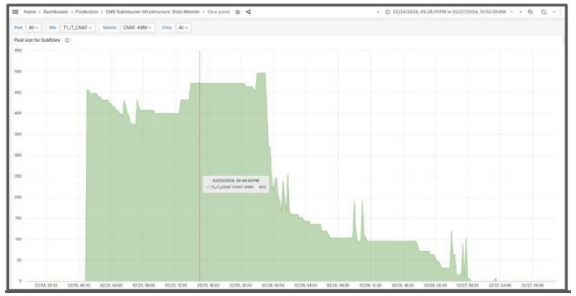


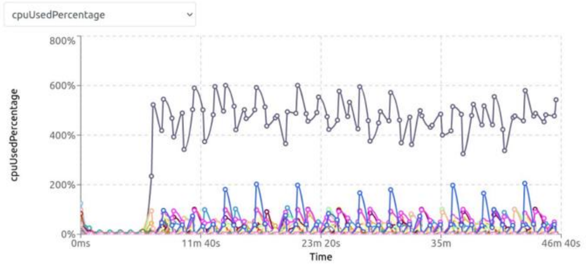
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

UC2.2.5 Physics validation of reconstruction code on ARM

Status by experiment

- ATLAS has been running jobs on ARM@CNAF for one year
  - Very good performance observed
    - steady use of available resources:
    - ~ on peak: 12% of ATLAS-dedicated resources at INFN-T1
- CMS: the ARM nodes at T1_IT_CNAF have been integrated
  - Technical validation fully done
  - Physics validation in progress
- ALICE preliminary tests: good balancing in using resources
  - still some instability in ARM builds
  - GRID submission @CNAF → validated





UC2.2.5 Physics validation of reconstruction code on ARM

- **KPI - Key Performance Indicator**

KPI ID	Description	Acceptance threshold	Status
KPI2.2.5.1	Software validation on ARM in the full GRID chain	50% (2/4 LHC experiments)	25%
KPI2.2.5.2	Presentation at conferences	≥ 2	100%
KPI2.2.5.3	Technical notes (in experiments and ICSC)	≥ 2	0%

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Open calls related to WP2

- D06-FaDER: Fast Data and Event Reduction: lowering data volumes in high-intensity experiments (Genova, FPGA)
  - PI: Carlo Schiavi, Supervisor: Alexis Pompili
- D11-DarkSieve: Ricerca di portali a settori Dark attraverSo l’identificazione di getti “boosted” originati da quark dark con tecniche di Machine Learning e nuove tecnologie ai futuri collisori adronici (Udine, UltrafastSim)
  - PI: Marina Cobal, Supervisor: Nicola de Filippis
- D05-BOODINI: BOOsting DIscoveries of New Interactions (Genova, UltrafastSim)
  - PI: Simone Marzani, Supervisor: Pietro Govoni
- D07-SPARCh: Simulation of Photons with Accelerated Ray-tracing for Cherenkov detectors (Genova, UltrafastSim GPU)
  - PI: Roberta Cardinale, Supervisor: Andrea Celentano
- D13-BOOST: Boosted Object and Oriented-Space Topologies from VBS@HL-LHC (Perugia, FPGA)
  - PI: Livio Fanò, Supervisor: Antonio Sidoti

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Summary

5 Flagship use cases as activity containers, overarching the whole activity of the WP.

Biweekly meetings, rotation of reports from each flagship, and talks on specific topics.

A large fraction of the KPIs has been completed.

Good interaction with WP4 and WP5.

We see cross contamination between experiments/WPs/institutions in action.