

## **PNNR - ICSC / Centro Nazionale “HPC, Big Data e Quantum Computing”**

**Kick-off meeting Spoke 2 - <https://agenda.infn.it/event/32786/> - 13/14 Ottobre 2022**

### **WP4 RECAP**

**S.Gennai (INFN-MiB) & A.Pompili (UniBA)**

**[Simone.Gennai@cern.ch](mailto:Simone.Gennai@cern.ch) / [alexis.pompili@ba.infn.it](mailto:alexis.pompili@ba.infn.it)**

## WP4: **missione**, **obiettivi**, “**milestone**”

### **Boosting the computational performance of Theoretical & Experimental Physics algorithms**

**4.1 Tools & guidelines** for developing and porting heterogeneous codes and algorithms on modern architectures

**4.2 Competence & Training Center** for heterogeneous computing

O4.1 : document best practises & SW tools for codes' development/porting to heterogeneous platforms (GPUs, FPGAs)

O4.2 : prepare and support the R&D testbed to offer multiple architectures; **optimize single-node performance**

O4.3 : organize training opportunities open to external users; trained personnel will help to boost the activities.

M9-15 : report on best practises for heterogeneous computing

M22-26 : first training opportunity; testbeds ready for users; user support in place

M25-36 : results from testbed & benchmarking activities

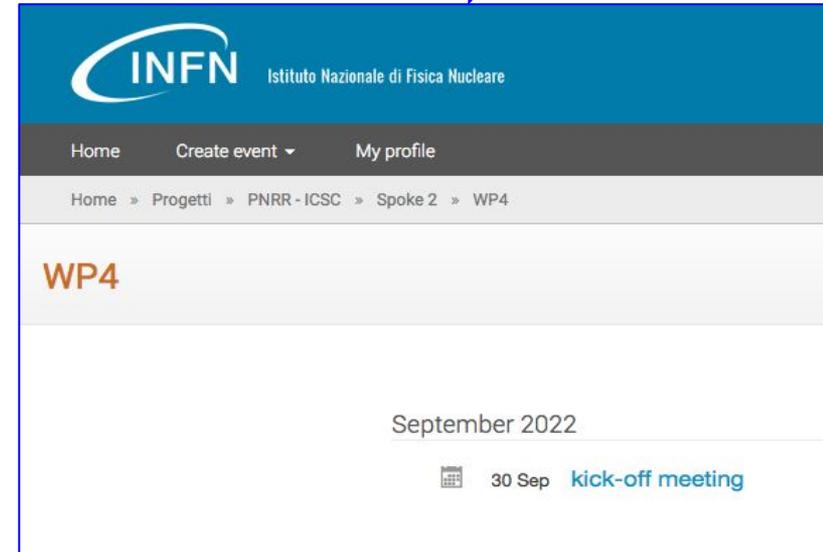
M36 : final report on technologies, training & support system. White Paper for use cases external to the CN.

## Riunioni periodiche del WP4 (periodicità mensile o bisettimanale a seconda della fase delle attività):

- Indico page : <https://agenda.infn.it/category/1780/> (agende dei meeting)
- Mailing list del WP4 : <https://lists.infn.it/sympa/info/cn1-spoke2-wp4-all>
- Mattermost channel: [https://cn1-spoke2-wp4.cloud.mattermost.com/signup\\_user\\_complete/?id=g5keafx9j7frtbqmyjcptxbqye](https://cn1-spoke2-wp4.cloud.mattermost.com/signup_user_complete/?id=g5keafx9j7frtbqmyjcptxbqye)

## Finora... **Attività esplorativa** :

- censire i progetti e use case in WP1-3 sinergici con WP4
- individuare (con Spoke e altri WP leaders) quelli che si prestano come benchmark progettuali ed il relativo necessario testbed
- capire il nuovo *person-power* necessario per portare avanti le attività; aiutare a preparare le call.
- cominciare (in bassa priorità) a pensare a come organizzare le attività formative aperte



# Recap della sessione di ieri pomeriggio

- Circa 20 connessi su zoom
  - compatibili con il numero di persone iscritte al canale MM e alla mailing list
- Abbiamo tentato di fare una overview del know how già acquisito nei vari gruppi
  - chiaramente non e' stato possibile avere una visione globale
  - ci saranno altre occasioni per discutere ulteriormente e coprire aree non ancora coperte

## Parallel WP4

ZOOM <https://us06web.zoom.us/j/88325395679?pwd=QTRDQXZ0bnRZCtBbWN0eUJDeDNJQT09>

Conveners: Alexis Pompili (Istituto Nazionale di Fisica Nucleare), Simone Gennai (MIB), Simone Gennai (Istituto Nazionale di Fisica Nucleare)

17:30 News

2022\_10\_13\_spoke...

17:35

### Porting to GPU, experience from Experimental-HEP

Speaker: Dr Tony Di Pilato

Tony\_Spoke2Perfor...

17:55

### Porting to GPU, experience from Theory-HEP

Speakers: Francesco Sanfilippo (Istituto Nazionale di Fisica Nucleare), Francesco Sanfilippo (INFN - Sezione Roma III)

2022-10-13.pdf

18:15

### FPGAs array per fast track reconstruction at trigger level

Speaker: Giovanni Punzi (Istituto Nazionale di Fisica Nucleare)

Punzi-ICSC-Spoke2...

18:35

### ML-INFN: cosa possiamo imparare ?

Speaker: Lucio Anderlini (Istituto Nazionale di Fisica Nucleare)



# Qualche considerazione **preliminare** al lavoro che ci aspetta

## ● GPU

- Ci sono già in corso diverse attività sia teoriche che sperimentali, al momento l'interesse maggiore e' lo sviluppo di framework platform agnostic
  - qui la maggiore esperienze viene dalla parte sperimentale, dove l'effort in questa direzione e' iniziato da tempo
  - serve rafforzare l'expertise italiana e fare formazione diffusa sul porting di algoritmi
  - al di la' del porting degli algoritmi c'e' poi da capire il fine tuning di diversi parametri per ottimizzare la memoria e l'esecuzione
    - ovviamente questo puo' essere molto project dependent

## ● FPGA

- Area con interesse crescente negli ultimi tempi, e **molto varia per quanto riguarda progetti e anche tipologia di HW**
  - attività principalmente lato sperimentale con forte liason a tematiche legate a trigger selection e reconstruction
- Probabilmente l'area dove c'e' meno expertise e quindi anche dove serve maggiormente la formazione di nuovo person-power
  - c'e' una conoscenza diffusa dei vari codici per la scrittura di firmware, manca forse la conoscenza per il setup di **cluster di FPGA** su media-grande scala (forse più indicato per WP5)
  - esistono vari testbed già funzionanti per alcuni progetti ma su scala ridotta, in altri casi si usano risorse messe a disposizione dai vari esperimenti (principalmente al CERN).

## ● Algoritmi ML

- Largamente in uso nella nostra comunità, con sempre maggiore interesse per deployment su acceleratori vari
- Per quanto riguarda il WP4, i punti di maggiore interesse potrebbero essere
  - ottimizzazione delle performance su GPU (eliminare i "colli di bottiglia" come il trasferimento dei dati o tuning di hyper-parameters
  - sviluppare know-how per il porting a FPGA (molti usano HLS4ML che e' in fase di completa riscrittura, ma sembrerebbe che nuovi tool siano in via di sviluppo)
- per quanto riguarda la formazione (hackaton, workshop, repository di esempi e mini tutorial) significativa e' l'esperienza già accumulata nell'ambito del **progetto ML\_INFN**: sarebbe opportuno approfittarne il piu' possibile evitando di duplicare lavoro già fatto.