



Advanced Machine Learning. Flash Simulation and bleeding edge applications

FlashSim: March status report

Lucio Anderlini

Istituto Nazionale di Fisica Nucleare, Sezione di Firenze

External Partner



Who we are

Staff members:

- Alessandro Bombini ^j, INFN
- Giuseppe Piparo ^l, INFN
- Maurizio Martinelli ^a, Università Milano Bicocca
- Simone Capelli ^a, Università Milano Bicocca
- Federica Maria Simone ⁱ, Politecnico di Bari
- Nicola De Filippis ⁱ, Politecnico di Bari
- Vieri Candelise ^h, Università di Trieste
- Giuseppe Della Ricca ^h, Università di Trieste
- Valentina Zaccolo ^k, Università di Trieste
- Mattia Faggin ^k, Università di Trieste
- Lorenzo Rinaldi ^e, Università di Bologna
- Piergiulio Lenzi ^g, Università di Firenze
- Vitaliano Ciulli ^g, Università di Firenze
- Sharam Rahatlou ^h, Università Roma 1
- Daniele del Re ^h, Università Roma 1
- Lorenzo Capriotti ^f, Università di Ferrara
- Francesco Conventi ^e, Università di Napoli
- Francesco Cirotto ^e, Università di Napoli

PhD students:

- Francesco Vaselli ^c, Scuola Normale Superiore di Pisa
- Matteo Barbetti ^b, Università di Firenze
- Muhammad Numan Anwar ^j, Politecnico di Bari
- Benedetta Camaiani ^g, Università di Firenze
- Alkis Papanastassiou ^g, Università di Firenze
- Antonio D'Avanzo ^e, Università di Napoli

External collaborators:

- Andrea Rizzi ^c, Università di Pisa



Update on infrastructure

The focus of last month has been on the deployment of the AI_INFN platform.

➤ <https://hub.ai.cloud.infn.it/>

Activities under the umbrella of the flagship are entitled to access GPU resources through the platform:

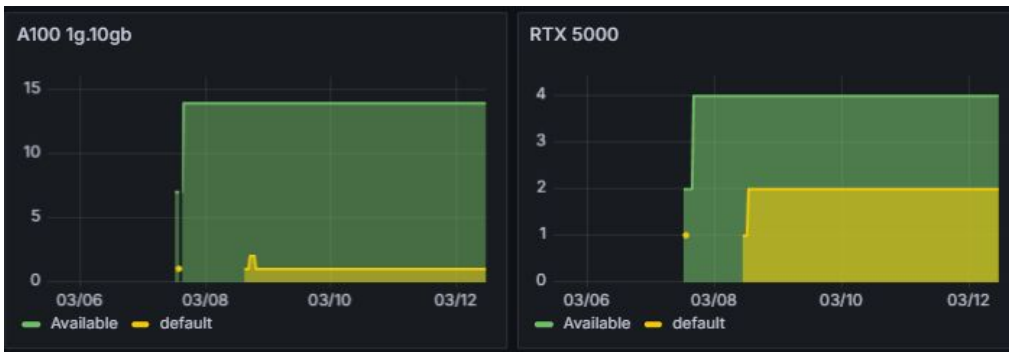
- request cloud credentials accessing <https://iam.cloud.infn.it>
- join the group “**ml-infn-platform**”

Resources are still somewhat limited, more will come with HPC bubbles.

Activity of the first few days



5 out of 8 users are connected to this flagship.



Infrastructure: developments

Done:

- public cvmfs was made available with CSI (Container Storage Interface)
- apptainer was made available in the default docker image to run docker-in-docker
- An integration for Snakemake workflows to kueue was drafted (it works, needs cleaning)

Ongoing:

- framed an activity with Spoke0 to enable the offloading of computations to remote, GPU-powered computing nodes using the InterLink solution
- discussions started to automate the deployment of the platform consistently with the technologies adopted and supported by INFN Cloud

KPIs

KPI ID	Description	Acceptance threshold	2024-02-13
KPI2.2.1.1	N_{MC} billion events obtained from ML-based simulation, as demonstrated by official links in experiments' simulation databases	$N_{MC} \geq 1$	1 M events (completed: 0.1%)
KPI2.2.1.2	N_{EXP} experiments have tested a machine-learning based simulation	$N_{EXP} \geq 2$	0 experiment (completed: 0%)
KPI2.2.1.3	Machine-learning use-cases tested in the context of the CN were presented at N_{CONF} international and national events	$N_{CONF} \geq 3$	3 use-cases (since Sept. '23) (completed: 100%)
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	$N_{UC} \geq 5$	5 use-cases (completed: 100%)

List of conferences for KPI2.2.1.3

- L.A., *Generative models at the LHC*, ALPACA workshop 2023, Trento
- B. Camaiani, *Example of adaptation domain in High Energy Physics*, XAI 2023, Milano
- A. Papanastassiou, *Anomaly detection with autoencoders for data quality monitoring in HEP*, XAI 2023, Milano

Abstract accepted (not counted in KPIs, yet.)

- M. Mazurek, *Lamarr: implementing the flash-simulation paradigm at LHCb*, ACAT 2024
- F. Simone, *Anomaly detection for data quality monitoring of the CMS detector*, AISSAI 2024
- *The flash-simulation of the LHCb experiment using the Lamarr framework*, EuCAIFCon 2024

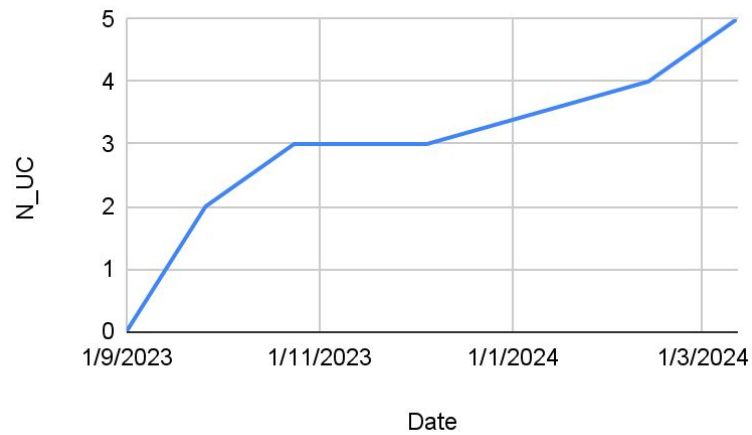
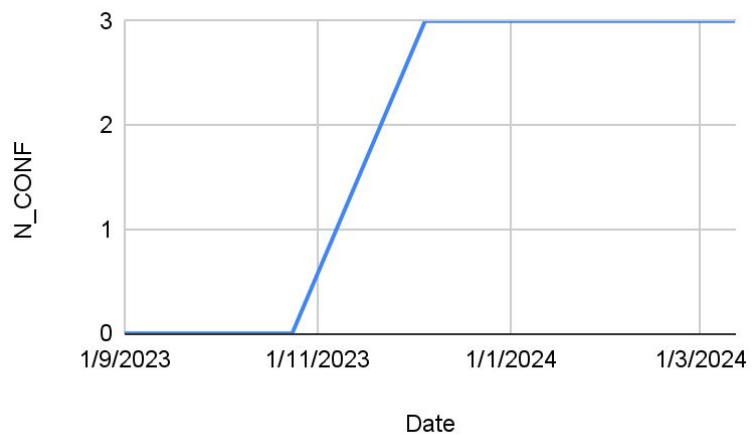
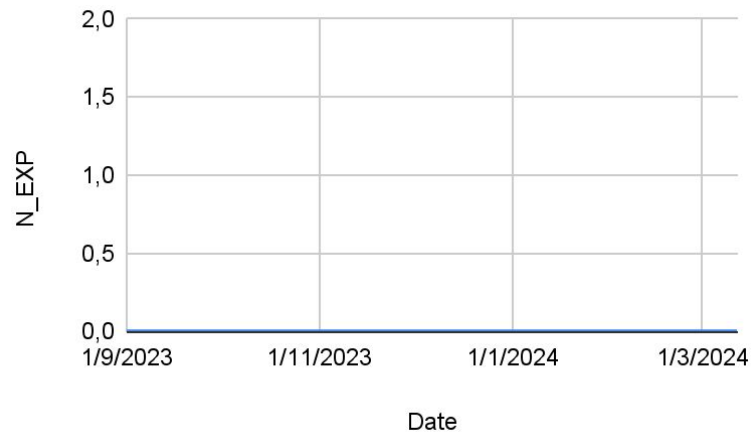
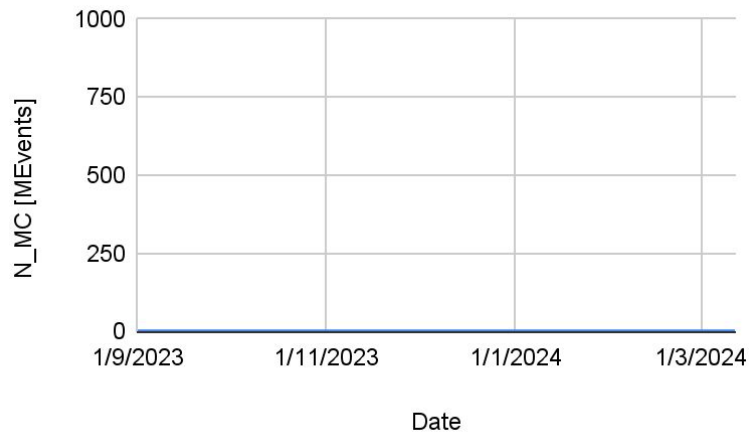
Abstract submitted

- A. Bombini, *Physics Informed Neural Networks for design optimisation of diamond particle detectors for charged particle fast-tracking at high luminosity hadron colliders*, PhMSML24
- M. Barbetti, *Generative models and seq2seq techniques for the flash-simulation of the LHCb experiment*, ICHEP 2024
- G. Piparo, *Reconstruction of multiple calorimetric clusters in the LHCf experiment with machine learning techniques*, ICHEP 2024

List of use-cases tested on the platform (100%)

- Lamarr, the ultra-fast simulation option for the LHCb experiment (tracking parametrizations)
- Lamarr, the ultra-fast simulation option for the LHCb experiment (particle identification and neutral reconstruction parametrizations)
- Theory-independent classifiers for the data analysis with the CMS experiment
- Machine-learning-based simulation of the response of resistive solid-state detector to the charge generated by a traversing minimum-ionizing particles
- Flavor tagging with GNN for the ATLAS experiment (new entry, INFN TS)

KPIs



On resources

The HPC bubbles requested by this flagship should be commissioned by fall 2024.

In the meanwhile:

- **Compute:**
 - ML_INFNO is providing GPUs for **interactive development** (no batch)
 - some of us is starting playing with ***Leonardo Booster*** for practicing **ML with batch jobs** (spoke, please clarify rules and procedures)
 - ML_INFNO plans for experimental support for batch jobs since early 2024, (but very tight schedule)
- **Storage:**
 - cache and ephemeral storage is available through INFNO Cloud (few TB)
 - “permanent” storage for data relevant for audit under discussion, for the time being, we are encouraged using minio.cloud.infn.it

Get in touch! The machinering is slowly starting to spin, since getting accustomed to the environment may require time, if you plan to use ICSC resources at some point, we suggest you start practicing with the temporary, resource-limited environment we have now.