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

Advanced Machine Learning. Flash Simulation and bleeding edge applications

FlashSim: *June status report*

with a focus on the Production strategy towards M12

Lucio Anderlini

Istituto Nazionale di Fisica Nucleare, Sezione di Firenze

External Partner

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

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

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} >= 1	2.3 M events (completed: 0.2%)
KPI2.2.1.2	N_{EXP} experiments have tested a machine-learning based simulation	N_{EXP} >= 2	3 experiment (completed: **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} >= 3	8 use-cases (since Sept. '23) (completed: **267%**)
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} >= 5	4 use-cases (completed: 80%)

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

List of conferences for KPI2.2.1.3

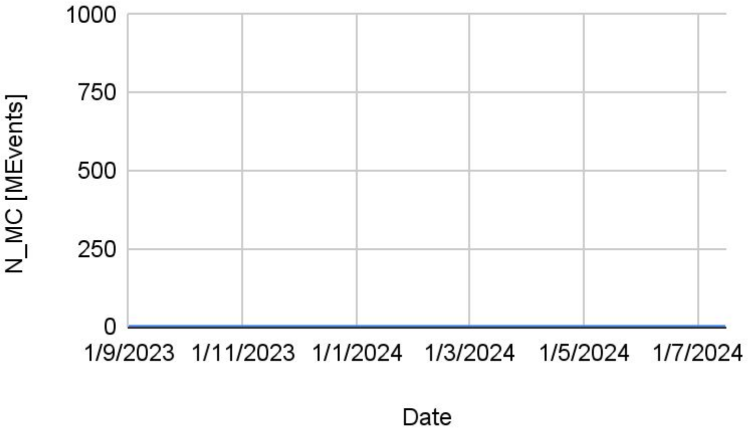
1. L.A., Generative models at the LHC, ALPACA workshop 2023, Trento
2. B. Camaiani, Example of adaptation domain in High Energy Physics, XAI 2023, Milano
3. A. Papanastassiou, “Anomaly detection with autoencoders for data quality monitoring in HEP”, XAI 2023, Milano
4. M. Mazurek (CERN), *Lamarr: implementing the flash-simulation paradigm at LHCb*, ACAT 2024
5. F. Simone, *Anomaly detection for data quality monitoring of the CMS detector*, AISSAI 2024
6. F. Corchia, *Tecniche computazionali avanzate per la simulazione veloce del calorimetro dell’esperimento ATLAS*, IFAE 2024
7. M. Barbetti, *The flash-simulation of the LHCb experiment using the Lamarr framework,* EuCAIFCon 2024
8. F. Vaselli, *FlashSim: an end-to-end fast simulation prototype using Normalizing Flow,* EuCAIFCon 2024

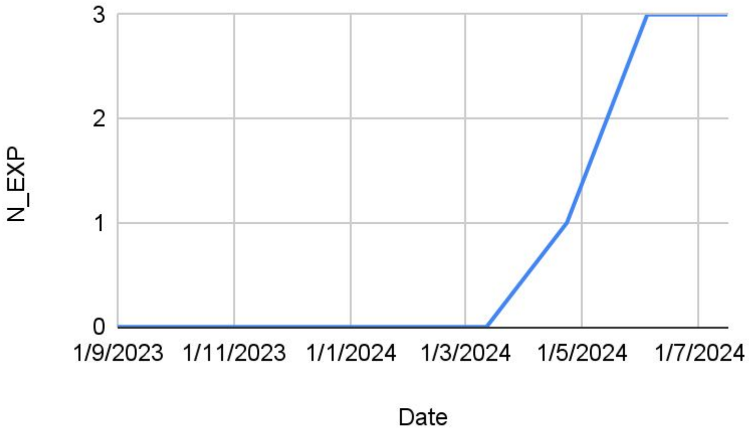
List of use-cases tested on the platform (⅘)

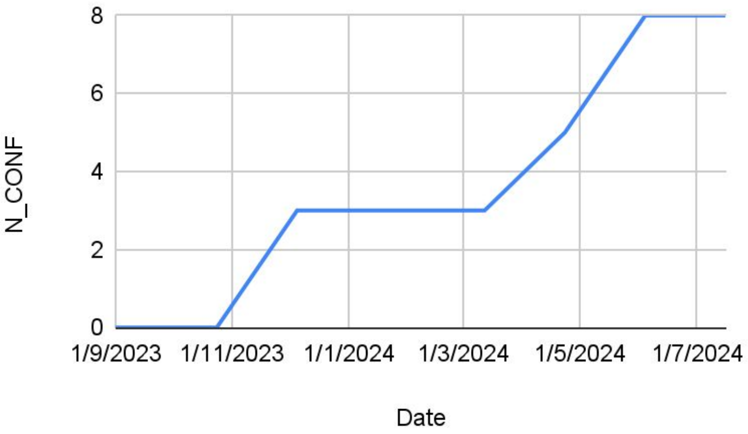
- Lamarr, the ultra-fast simulation option for the LHCb experiment (tracking parametrizations)
- Lamarr, the ultra-fast simulation option for the LHCb experiment (particle identification and neutral reconstruction parametrizations)
- Theory-independent classifiers for the data analysis with the CMS experiment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

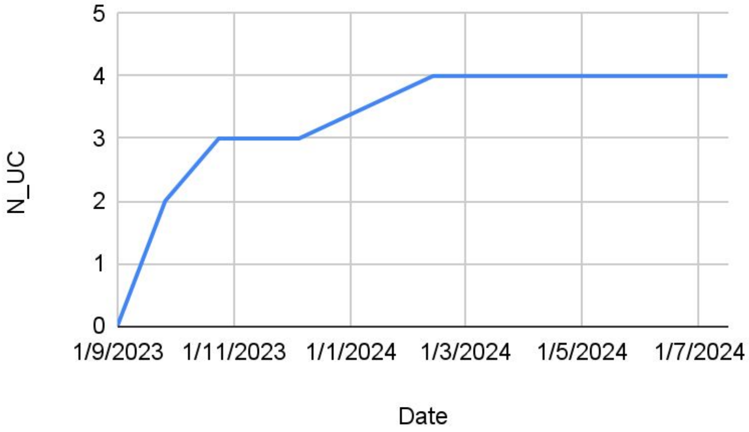
+ *Preliminary discussion with Muhammad Numan Anwar to bring HPO in the Cloud platform*

KPIs


1000
750
500
250
0
N_MC [MEvents]
1/9/2023
1/11/2023
1/1/2024
1/3/2024
1/5/2024
1/7/2024
Date


3
2
1
0
N_EXP
1/9/2023
1/11/2023
1/1/2024
1/3/2024
1/5/2024
1/7/2024
Date

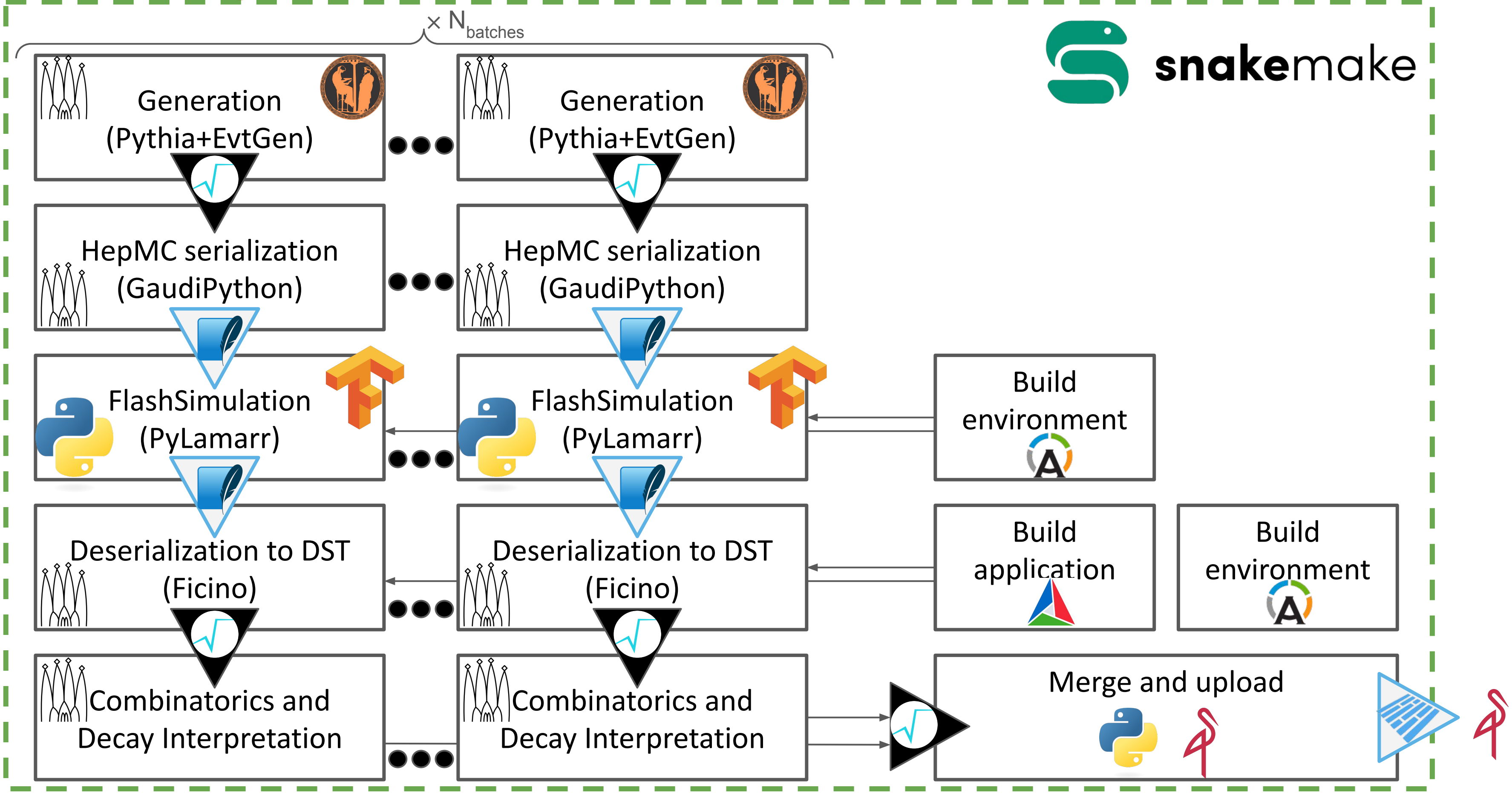

8
6
4
2
0
N_CONF
1/9/2023
1/11/2023
1/1/2024
1/3/2024
1/5/2024
1/7/2024
Date


5
4
3
2
1
0
N_UC
1/9/2023
1/11/2023
1/1/2024
1/3/2024
1/5/2024
1/7/2024
Date

Lamarr productions

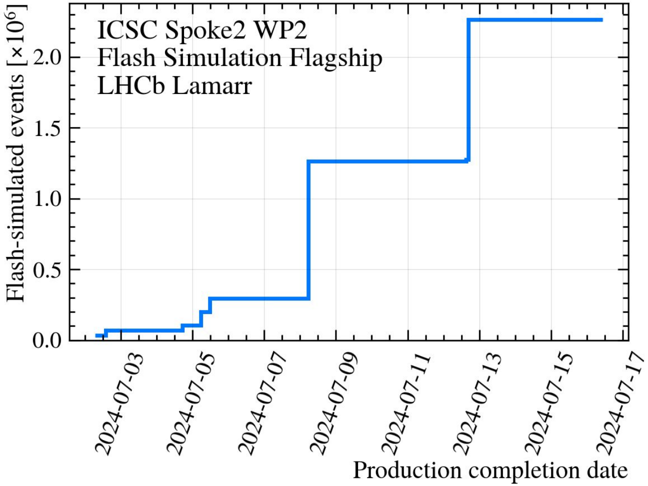
a.k.a. how to achieve the 1B events KPI in 12 months

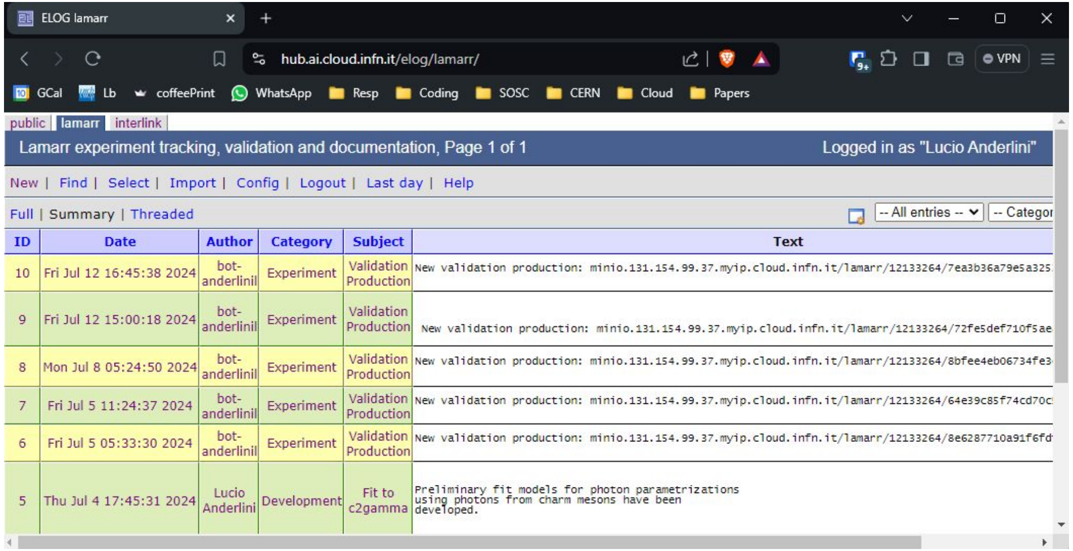
Recalling the production strategy


× N_batches
snakemake
Generation (Pythia+EvtGen)
HepMC serialization (GaudiPython)
FlashSimulation (PyLamarr)
Deserialization to DST (Ficino)
Combinatorics and Decay Interpretation
Generation (Pythia+EvtGen)
HepMC serialization (GaudiPython)
FlashSimulation (PyLamarr)
Deserialization to DST (Ficino)
Combinatorics and Decay Interpretation
Build environment
Build application
Build environment
Merge and upload

Measuring the produced number of events

To keep track of the productions, we deployed an instance of PSI ELOG in the AI_INFN platform, integrating its authentication procedure with Indigo IAM (original contribution).





Lamarr experiment tracking, validation and documentation, Page 1 of 1 | Logged in as "Lucio Anderlini"

New | Find | Select | Import | Config | Logout | Last day | Help

Full | Summary | Threaded | -- All entries -- | -- Categor

ID	Date	Author	Category	Subject	Text
10	Fri Jul 12 16:45:38 2024	bot-anderlinil	Experiment	Validation Production	New validation production: minio.131.154.99.37.myip.cloud.infn.it/lamarr/12133264/7ea3b36a79e5a325
9	Fri Jul 12 15:00:18 2024	bot-anderlinil	Experiment	Validation Production	New validation production: minio.131.154.99.37.myip.cloud.infn.it/lamarr/12133264/72fe5def710f5ae
8	Mon Jul 8 05:24:50 2024	bot-anderlinil	Experiment	Validation Production	New validation production: minio.131.154.99.37.myip.cloud.infn.it/lamarr/12133264/8bfee4eb06734fe3
7	Fri Jul 5 11:24:37 2024	bot-anderlinil	Experiment	Validation Production	New validation production: minio.131.154.99.37.myip.cloud.infn.it/lamarr/12133264/64e39c85f74cd70c
6	Fri Jul 5 05:33:30 2024	bot-anderlinil	Experiment	Validation Production	New validation production: minio.131.154.99.37.myip.cloud.infn.it/lamarr/12133264/8e6287710a91f6fd
5	Thu Jul 4 17:45:31 2024	Lucio Anderlini	Development	Fit to c2gamma	Preliminary fit models for photon parametrizations using photons from charm mesons have been developed.

We patched the Python client to use the user token in JupyterLab to connect to the elog in Python scripts.

Completed productions are **logged in the ELOG** with a snapshot of the configuration used and some statistics.

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

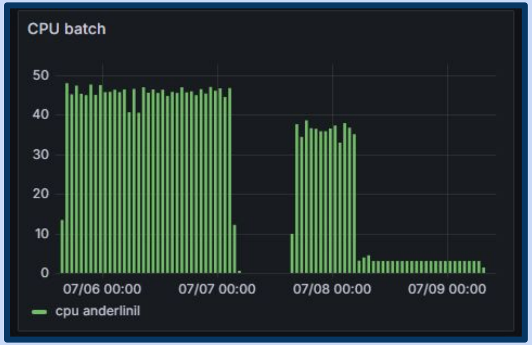
Resources

Pythia8 (full event)

Generates the whole proton-proton collision event, with pileup and spill-over.
Then processes all particles with Lamarr and Bender to produce nTuples.

1M events (on 50 parallel jobs) require:

- O(48h) × 50 CPUs
- 0.8 TB of buffer in S3.

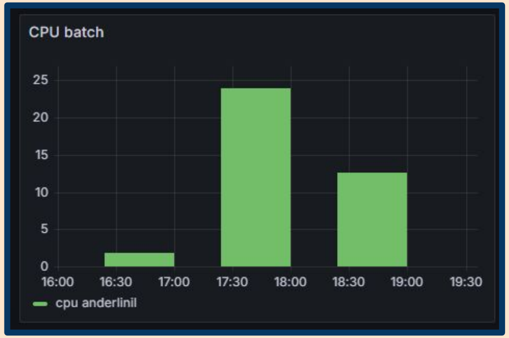


Particle Gun (signal-only)

Generates only the heavy hadron decay.
Then processes particles with Lamarr and Bender to produce nTuples.
Less tested than Pythia8 productions

1M events (on **up to** 50 parallel jobs) require:

- O(1h), *limited by submission latency*
- 4 GB of buffer in S3



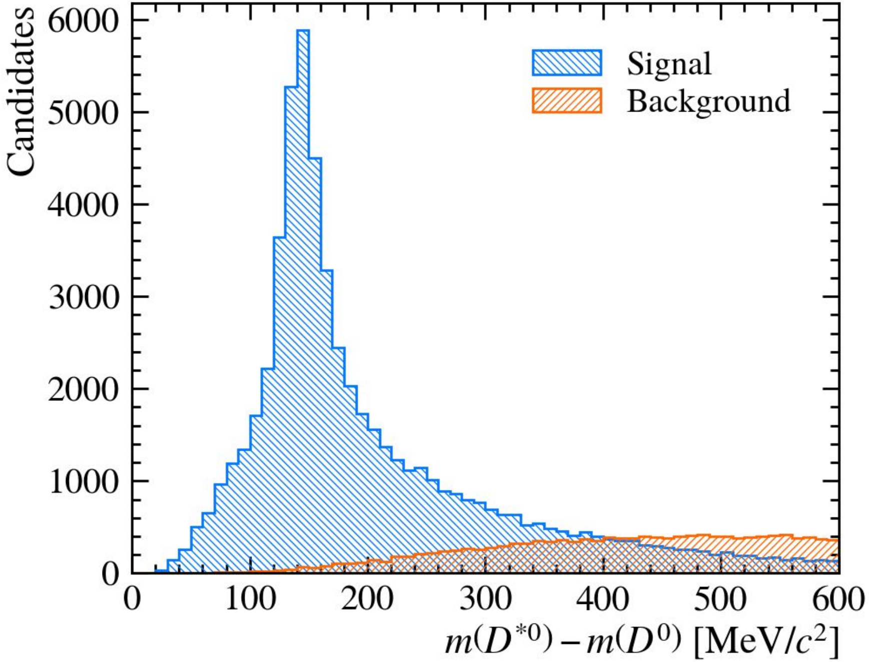
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Can we make everything with ParticleGun?

Unfortunately no.

Studying the particle-to-particle correlation is the frontier of Flash Simulation research and require **simulating non-signal particles to study how models predict their effect on the signal.**

The good news is we don't really need 1B events for those studies.



Additional CPU-saving strategies

Most of our simulations are used to validate parametrizations and software in Lamarr.

Most often, there is no need to re-generate the events with Pythia.
We can reuse the Pythia events saving up to 50% of the processing time.

This has an impact on S3 storage (now using 4 TB of AI_INFN Ceph storage, a ×2 is possible).

With **ParticleGun**, we are limited by the *job submission system*:

$$T_{\text{submission}} > T_{\text{job}} / n_{\text{CPUs}}$$

$T_{\text{submission}}$ is “large by design” as too high submission frequency may disturb the Kubernetes API server of the AI_INFN Platform (running in production), but it was never tuned.
T_{job} can be made larger increasing the event-batch size (in Pythia it is 1000)

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

Accessing remote resources

We remain confident that we will be able to access ICSC resources via InterLink.

A significant part of the flagship is to demonstrate we can use those resources transparently. InterLink has been validated against multi-cloud resources and significant effort is ongoing to keep it updated and running (*at least somewhere*).

However,

- The **small number of remote cores** we can access today does not justify that effort as it does not provide a viable option to scale up the Lamarr productions
- Validation of the Lamarr workflow with **batch-system backends** is lagging (realistically, it will not happen before late October)
- It is not clear **when** and **how** **HPC bubble resources** will be provisioned

Conclusion

With the resources available as of today,
it would take **2000 days** to produce **1B events with Pythia + Lamarr**.

With low-cost optimizations and the promised INFN-Cloud resources this might be pushed to down to 400-500 days. Which would remains challenging.

A (possibly large) fraction of the 1B events can be generated with ParticleGun.
With the resources available today,
it would take **45 days** to produce **1B events with Particle Gun + Lamarr**.

Additional development is needed to make Particle Gun useful.
It is being prioritized over other developments.

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

Requests for the validation part

Resource	Full Request	Strictly required for KPI 1 (Full-Pythia option)
CPU on INFN Cloud	2 M CPU hours	2.4 M CPU hours*
GPU on INFN Cloud	4 H200 for 18 months	0
GPU on Leonardo Booster via InterLink	10000 hours	0
Storage	25 TB	10 TB

Preliminary

- 0.5 M hours from opportunistic borrowing from AI_INFN Platform