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 Lamarr and offloading

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

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	1 M events (completed: 0.1%)
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%)

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

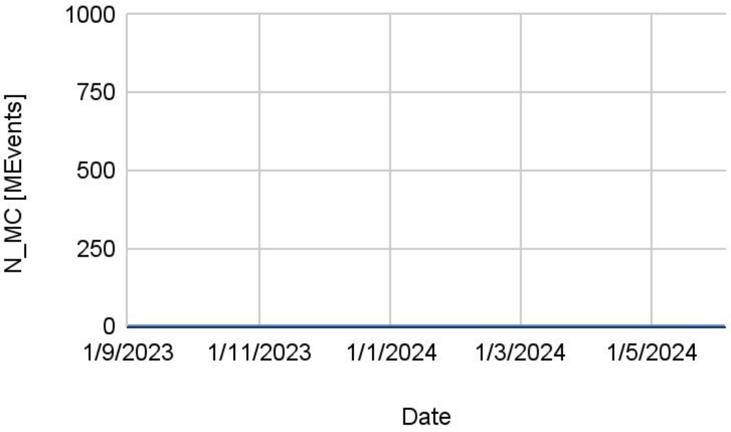
N_EXP += 2

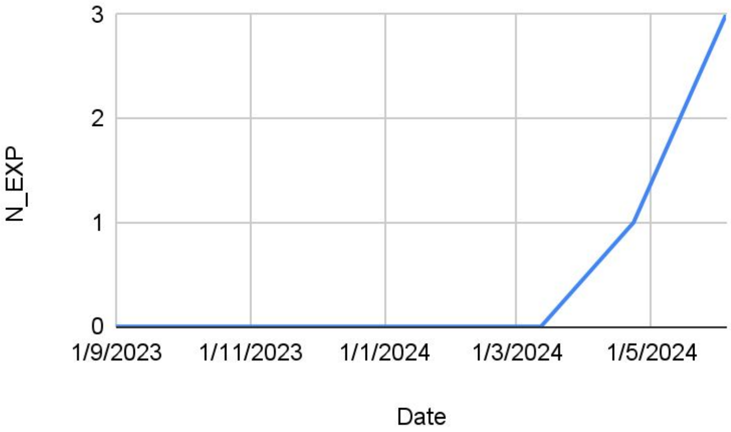
List of use-cases tested on the platform (⅘)

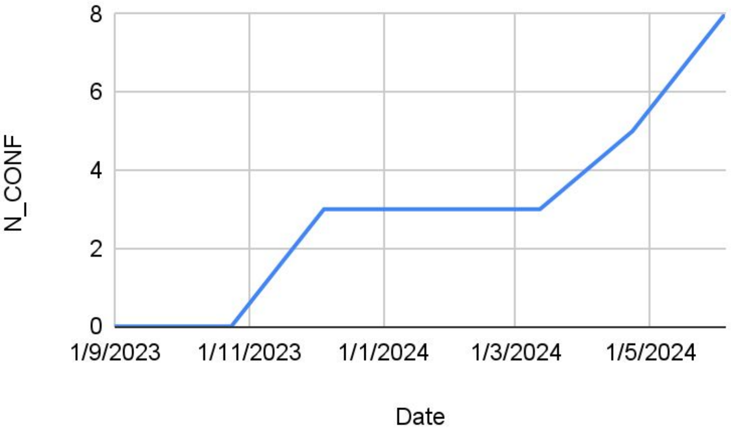
- Lamarr, the ultra-fast simulation option for the LHCb experiment (tracking parametrizations)
- Lamarr, the ultra-fast simulation option for the LHCb experiment (particle identification and neutral reconstruction parametrizations)
- Theory-independent classifiers for the data analysis with the CMS experiment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

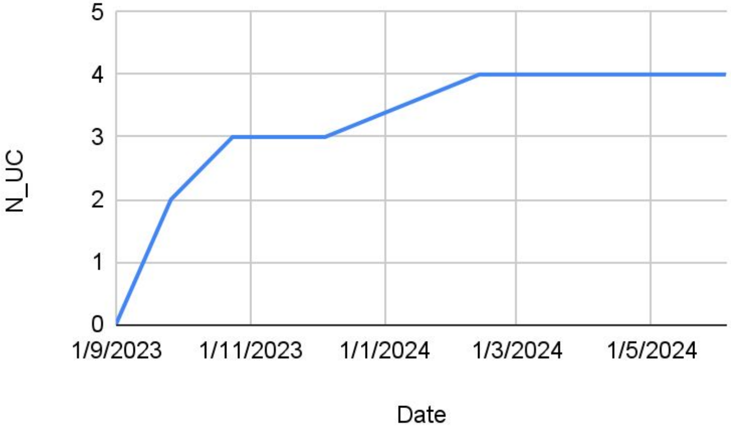
+ *Preliminary discussion with Muhammad Numan Anwar to bring HPO in the Cloud platform*

KPIs


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


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


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


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

Lamarr validation workflow

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

Introduction

Lamarr is the *Flash Simulation* option for the LHCb experiment.

Its development involves a complex workflows for training and validation that challenges the Cloud infrastructure.

It was selected as case-study to *enable machine learning* studies with Cloud-based resources.

Features of the workflow:

- ✓ uses LHCb software via `cvmfs`
- ✓ requires heterogeneous resources (GPUs)
- ✓ combine HTC steps (*Pythia*) with HPC steps (Tensorflow)
- ✓ combine steps with inconsistent software dependencies

Snakemake as a workflow manager

While resource greedy, we intend FlashSim as managed by analysts rather than centrally. Let’s call it a *quasi-interactive simulation*.

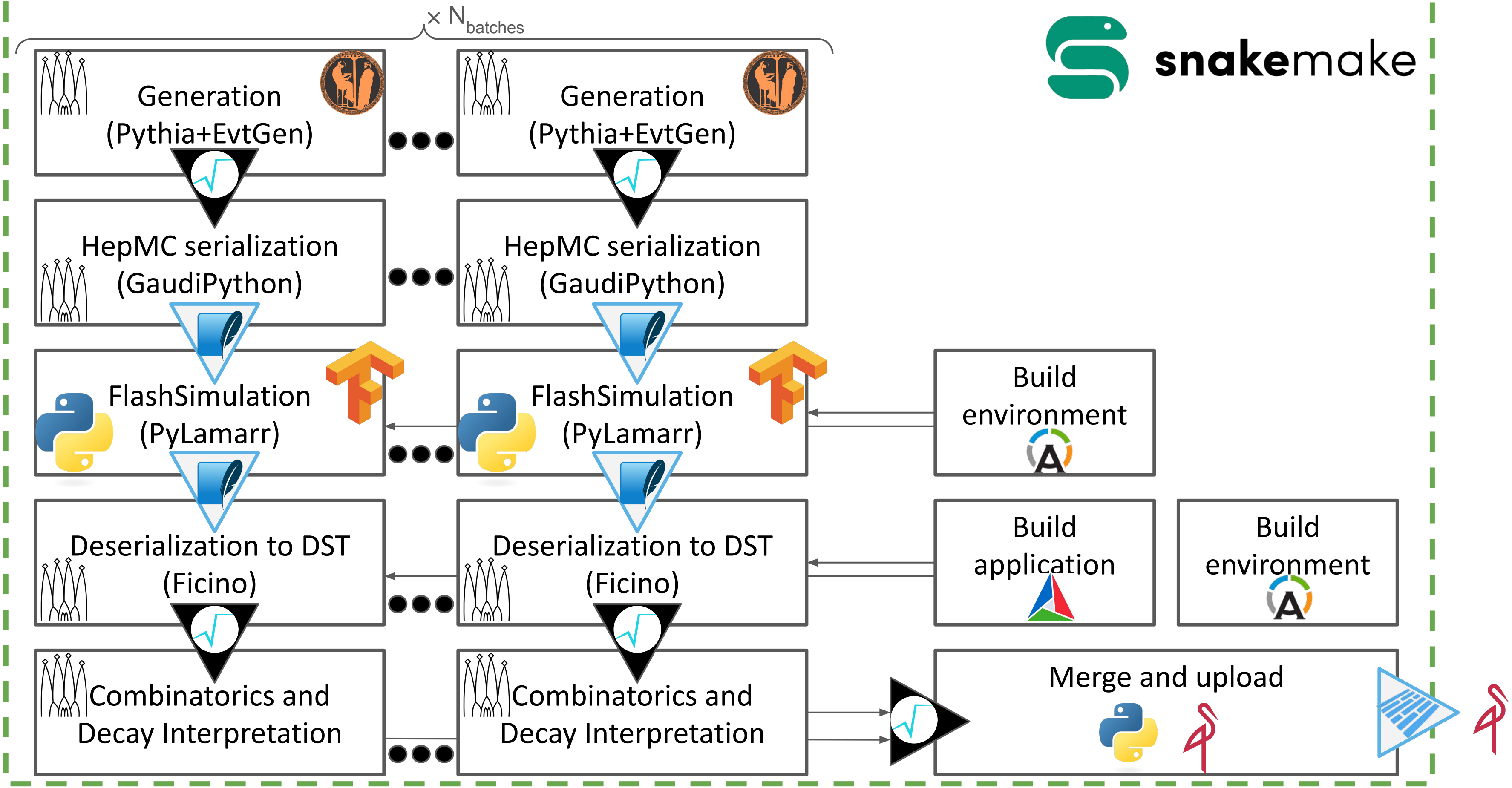
We selected Snakemake as a workflow manager:

- can run everywhere from a laptop to CINECA Leonardo
- supported by the reana initiative (docs.reana.io/)
- integrated with multiple cluster and ***cloud resources*** (through plugins)
- supporting both **standard and custom resources** (GPU, sites, …)
- support both conda and **apptainer** for environment management
- can define complex **DAG workflow** with caching of intermediate results

*Main challenge: works best with a **distributed filesystem**.*

Lamarr Validation workflow

[github.com/LamarrSim/validation-workflow (still private)]

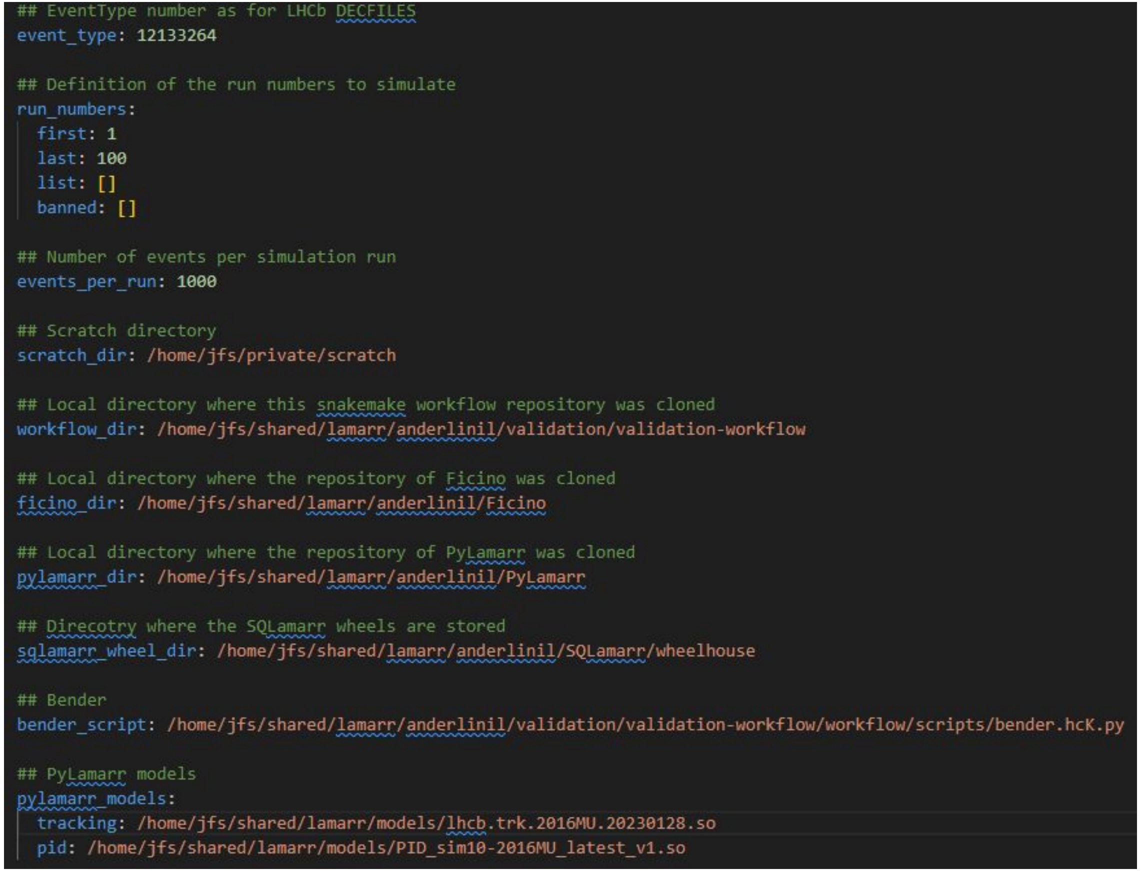


× $N_{batches}$
snakemake
Generation (Pythia+EvtGen)
HepMC serialization (GaudiPython)
FlashSimulation (PyLamarr)
Deserialization to DST (Ficino)
Combinatorics and Decay Interpretation
Generation (Pythia+EvtGen)
HepMC serialization (GaudiPython)
FlashSimulation (PyLamarr)
Deserialization to DST (Ficino)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

Lamarr Validation workflow configuration

The configuration of the workflow is described in YAML files defining:

- what to generate
  (for event_type see the DECFILE package)
- how to split events in batches
- how to simulate
- what to put in the nTuples



```yaml
## EventType number as for LHCb DECFILES
event_type: 12133264

## Definition of the run numbers to simulate
run_numbers:
  first: 1
  last: 100
  list: []
  banned: []

## Number of events per simulation run
events_per_run: 1000

## Scratch directory
scratch_dir: /home/jfs/private/scratch

## Local directory where this snakemake workflow repository was cloned
workflow_dir: /home/jfs/shared/lamarr/anderlinil/validation/validation-workflow

## Local directory where the repository of Ficino was cloned
ficino_dir: /home/jfs/shared/lamarr/anderlinil/Ficino

## Local directory where the repository of PyLamarr was cloned
pylamarr_dir: /home/jfs/shared/lamarr/anderlinil/PyLamarr

## Direcotry where the SQLamarr wheels are stored
sqlamarr_wheel_dir: /home/jfs/shared/lamarr/anderlinil/SQLamarr/wheelhouse

## Bender
bender_script: /home/jfs/shared/lamarr/anderlinil/validation/validation-workflow/workflow/scripts/bender.hcK.py

## PyLamarr models
pylamarr_models:
  tracking: /home/jfs/shared/lamarr/models/lhcb.trk.2016MU.20230128.so
  pid: /home/jfs/shared/lamarr/models/PID_sim10-2016MU_latest_v1.so
```

Executing the workflow in the cloud

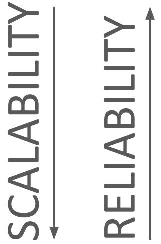
We are interfacing Snakemake to the AI_INFN Platform (hub.ai.cloud.infn.it).

AI_INFN
Artificial Intelligence technologies for INFN research

jupyterhub

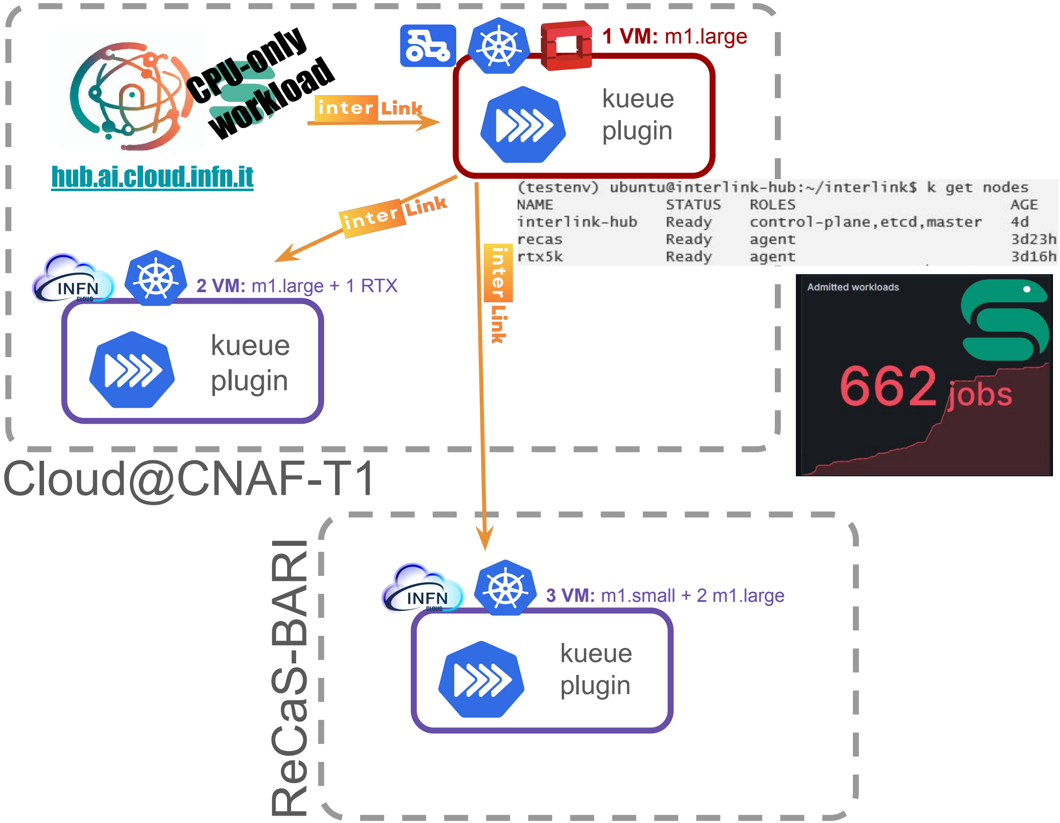
Snakemake steps can be executed:

- locally in the user’s JupyterLab
- in the cluster, using resources destinated to interactive usage opportunistically
- in the Cloud, via interLink offloading

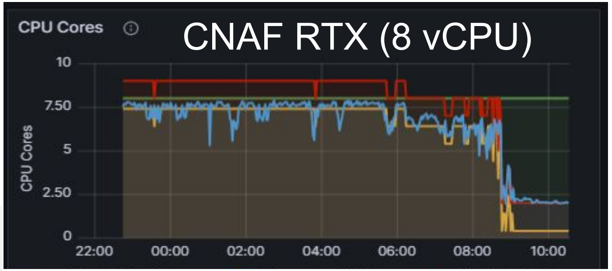


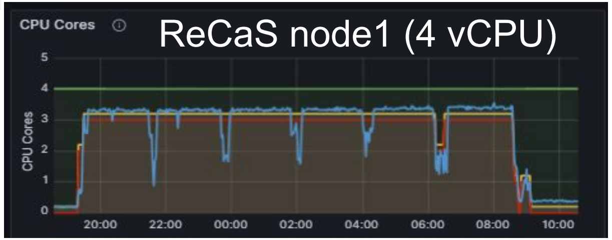
Filesystem distribution powered by JuiceFS

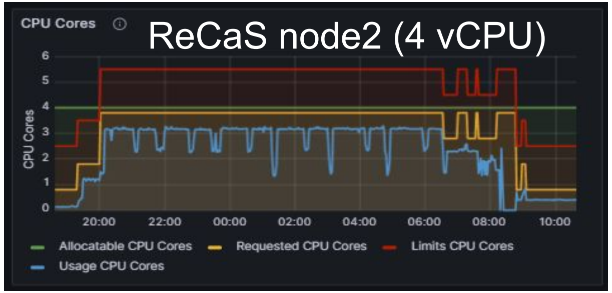
First attempt of running on multiple sites



```
(testenv) ubuntu@interlink-hub:~/interlink$ k get nodes
NAME            STATUS   ROLES                        AGE
interlink-hub   Ready    control-plane,etcd,master    4d
recas           Ready    agent                        3d23h
rtx5k           Ready    agent                        3d16h
```



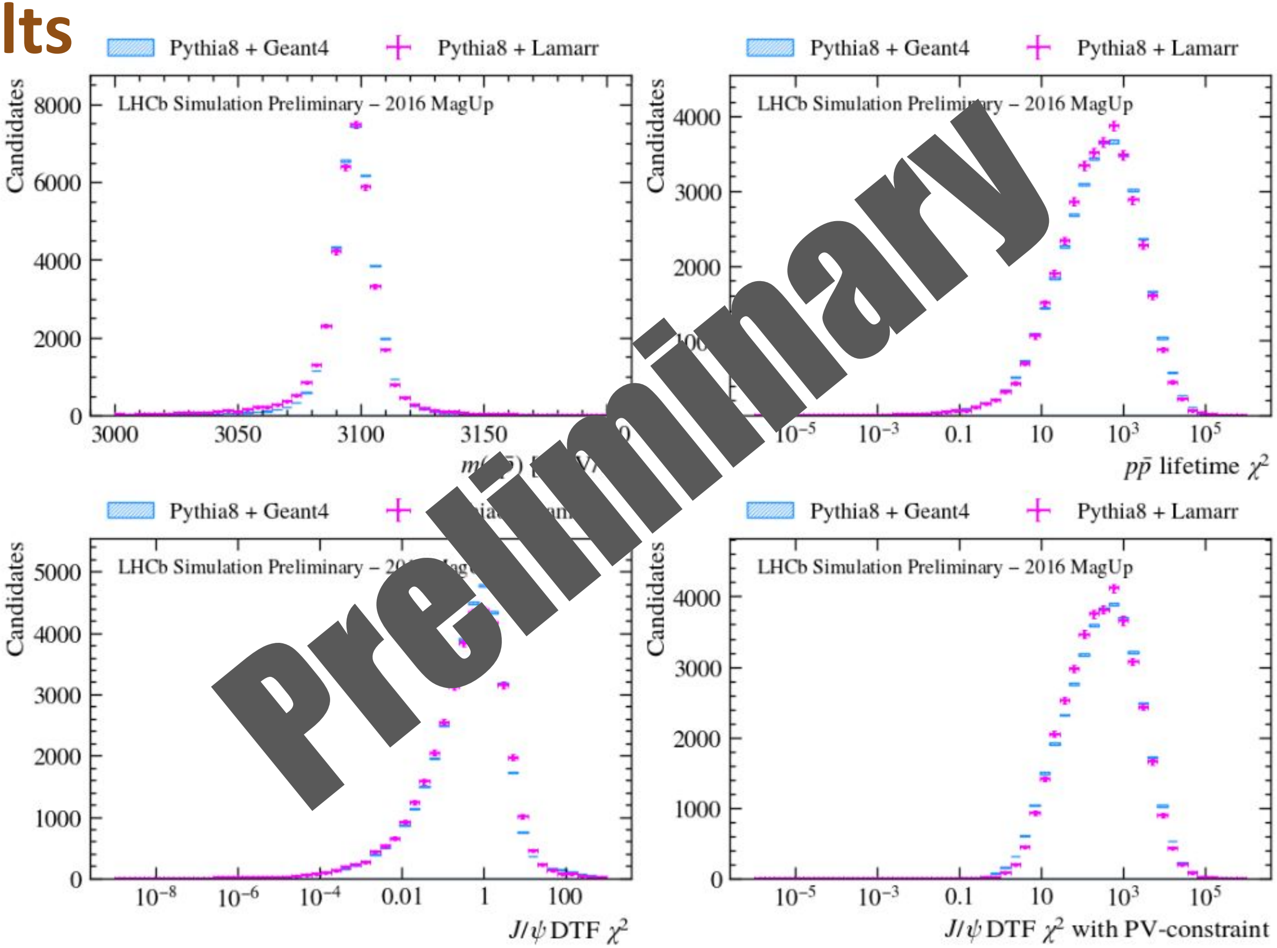




Preliminary results

We are looking at a B decay to a final state including a J/$\psi \longrightarrow$ pp̄.

Here are the very preliminary comparison plots.



Conclusion

- **Deploying Lamarr on Cloud resources is a ice-breaker for several other machine-learning workflows** combining CERN C++ software and machine learning;
- We have a first prototype of the whole workflow running through interLink on multi-cloud resources;
- Soon (this evening?) **other Cloud providers** will join, starting with **CloudVeneto**;
- At some point, interLink offloading towards **CINECA Leonardo** will be possible.

- Too many things are still *ad-hoc* for Lamarr, need work on **generalization** and **documentation**
- Upcoming feature: need work to support mixed workflows with the same step running both locally and in offloading