The AI_INFN initiative

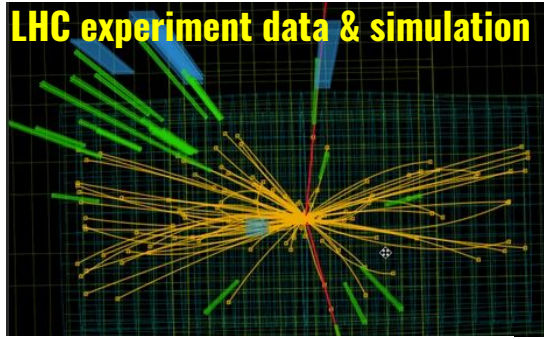
Daniele Spiga *on behalf of the proponents*

AI_INFN
Artificial Intelligence technologies for INFN research

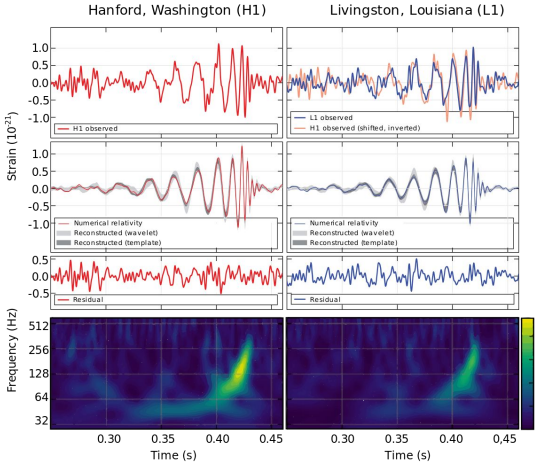
Machine Learning Technologies for INFN

Most of the experiments and initiatives produce, analyse or process digital data.

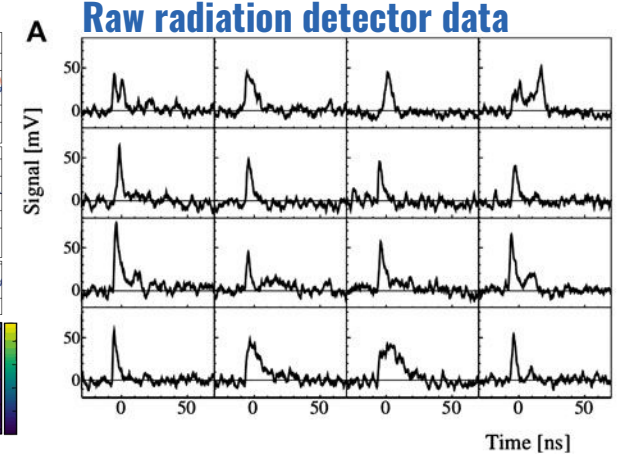
Enthusiasm on the modern data processing technologies!

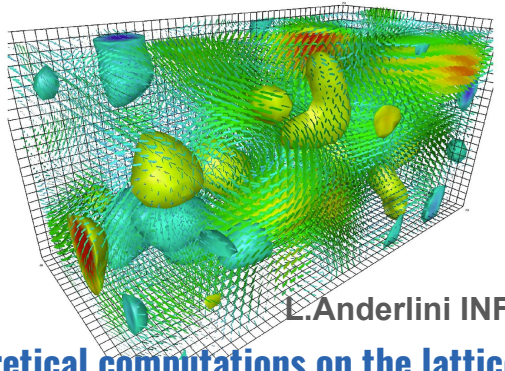


Gravitational wave detection



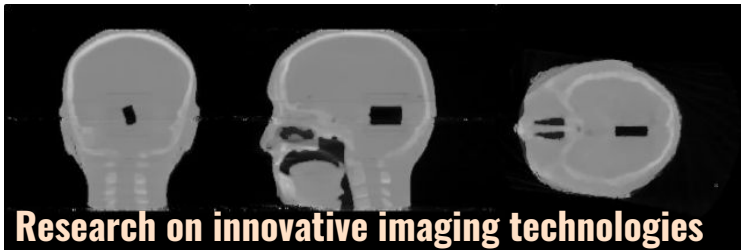
Raw radiation detector data





L.Anderlini INFN-F

Theoretical computations on the lattice

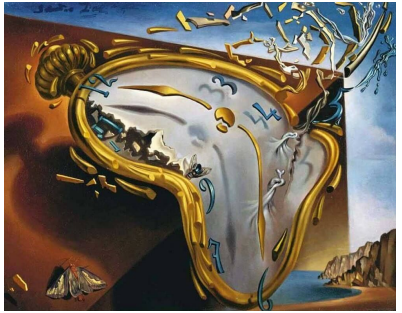


State of the art and ML_INFN

The ML_INFN initiative was proposed in 2020 at the dawn of the **INFN Cloud initiative**.

- Commissioned at CNAF a farm with capable of handling **up to 48 simultaneous user** sessions accessing data-center level GPU resources; served via INFN Cloud.
- **Designed and organized 4 educational events** targetting two levels of proficiency (beginner and advanced); highly oriented to discuss the code in small teams.
- Collected and organized examples from success stories of applications of machine learning at research topics in a dedicated web page: **The ML_INFN Knowledge Base**

Four years after, the landscape has changed



L.Anderlini INFN-Fl

- INFN is leading **the ICSC and TeRABIT initiatives**, funded on PNRR resources, exporting the INFN Cloud model to a wider community and wealth of GPU resources, with the name **DataCloud**.
- New models and approaches (*Transformers, Graph Neural Networks, Physics-Informed Neural Networks, Large Language Models, Differentiable Programming…*) have drastically **widened the application range of ML**
- Most Academic Degrees in Physics feature (at least) **entry-level courses on ML** for data analysis, many entry-level courses provided by *Ufficio Formazione*
- New hardware and computing technologies are arising as "*specialized accelerators*" for performing machine learning at scale: ***Quantum Computing* and *FPGAs***.

It's time to renew ML_INFN to make it ready for the upcoming challenges!

WP1 Infrastructure and Resource Provisioning

Lots of resources coming from ICSC and TeRABIT?

→ Less "*pressure for being in production*" on our farm

→ Opportunity for contributing to the provisioning model

Focus shifts towards R&D on the provisioning model, with a systemic view to ease ML workloads.

Needs for an updated and well maintained farm.

Open Science and Advanced Education **WP2**

What will be added-value in our *hackathons*?

→ ML_INFN has attracted a community of world-leading experts in the application of ML to research in physics

→ We canore ambitious in the target of our *hackathons,* letting experts to discuss their code

Focus shift towards *Advanced Hackathon Workshops.*

AI_INFN
Artificial Intelligence technologies for INFN research

Scientific use cases

Applications to scientific research remains central.

To develop the tools for making it easier to do machine learning for INFN researchers,
we need them to use to tools and provide feedback.

WP3 ***How?*** **User support and community engagement**

ML on FPGA and Quantum Computers

New hardware will change the landscape of computing.

Deploying ML algorithms of **FPGAs** enables fixed-latency, low-energy inference.

Quantum Computing will enable extremely fast computations of specialized, possibly trained, algorithms.

WP4

L.Anderlini INFN-FI

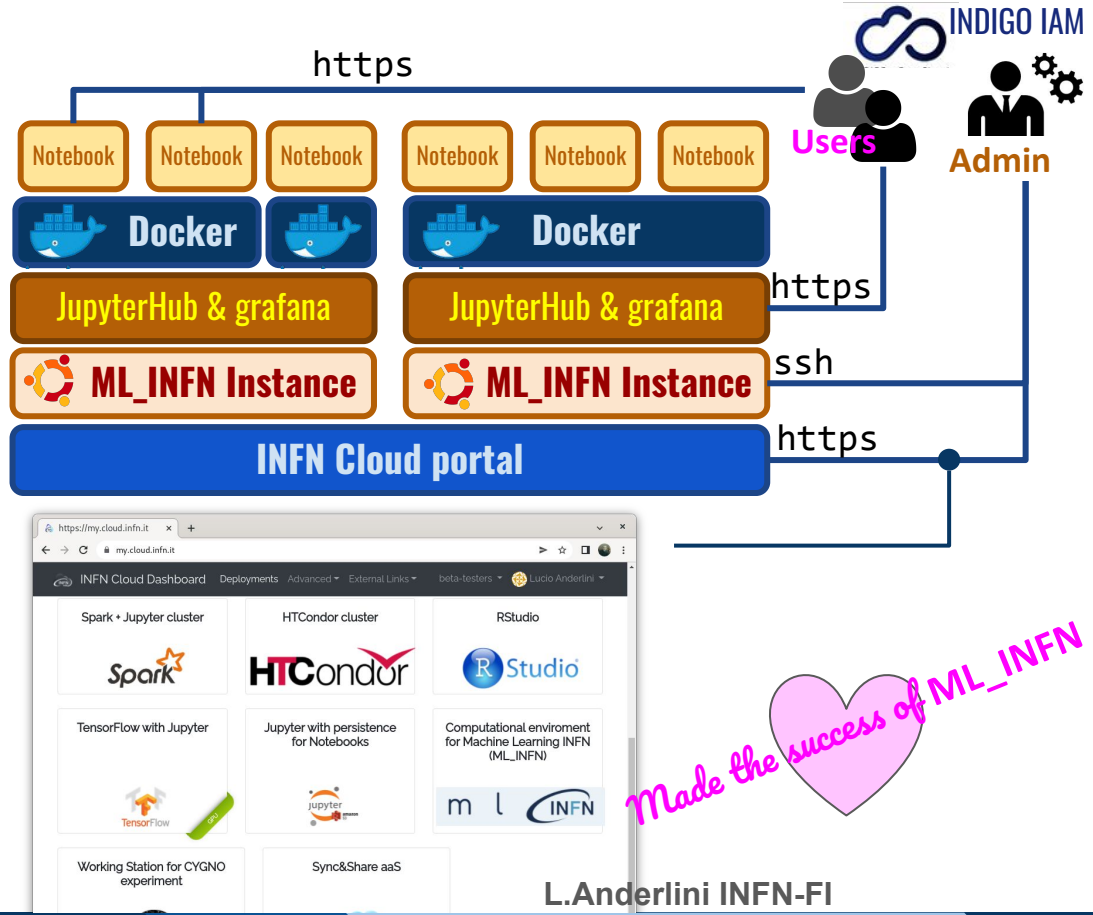
The provisioning model: ML_INFN version

Each project gets its own Virtual Machine

At the end of the project, the VM is destroyed, the GPU is freed for other users/projects, data in the filesystem is lost.

Resources are **guaranteed to the project**

Inefficient and **too many admins.**



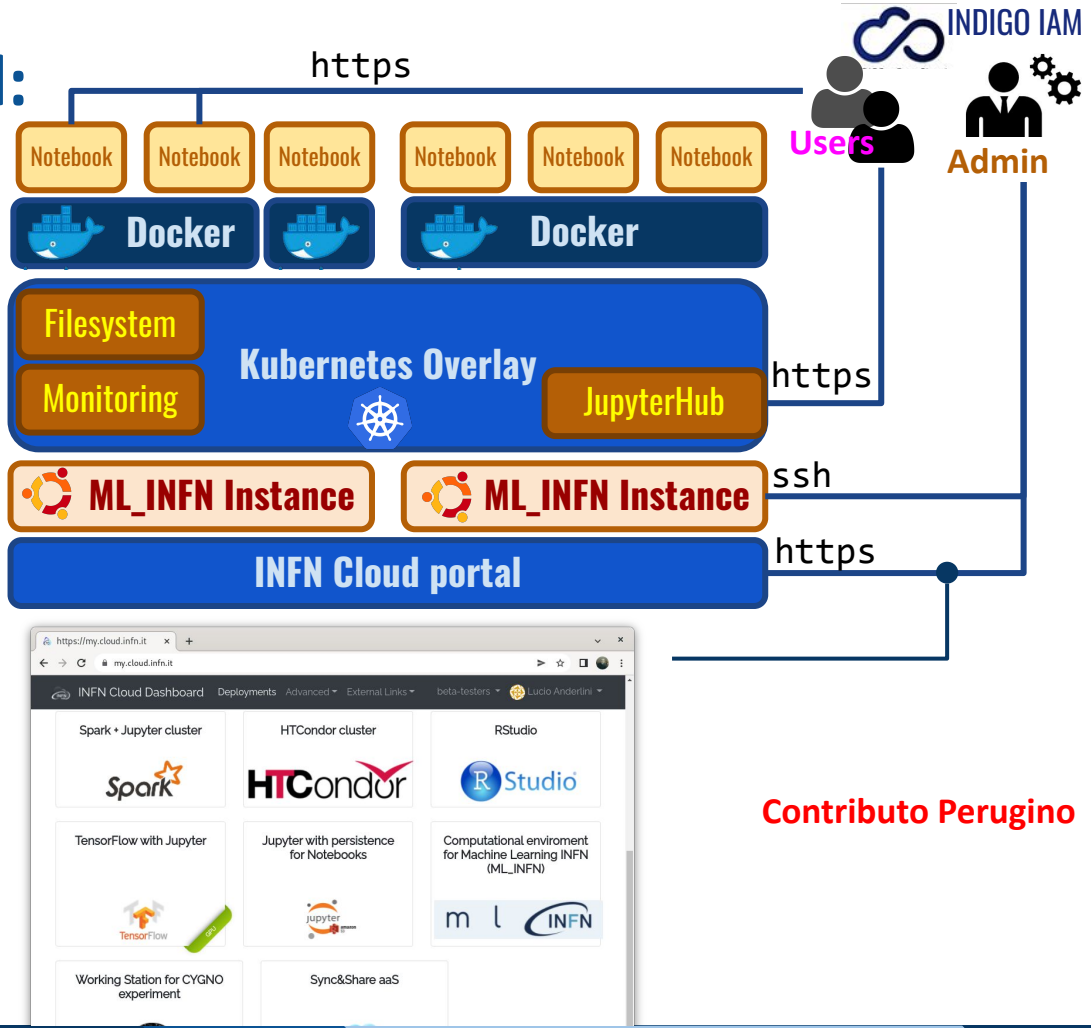
The provisioning model: AI_INFN version

An additional abstract, elastic overlay is added on top of multiple VMs.

Adding and removing machines enables **scaling based on demand**.

Filesystem is persistent at platform-level: GPUs can be re-assigned without data loss.

Guarantee access to resources will require custom policies.



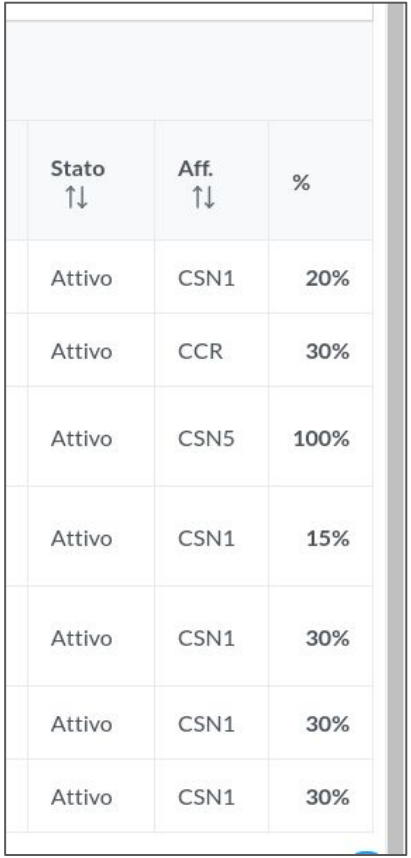
Contributo Perugino

Anagrafica

25 ricercatori e 12 tecnologi
(+ sinergie importanti con ICSC, TeRABIT e FAIR)

Unità coinvolte e Resp. Locali

- BA - Alfonso Monaco
- BO - Daniele Bonacorsi
- CNAF - Stefano Dal Pra
- FE - Enrico Calore
- FI - Lucio Anderlini (Responsabile Nazionale)
- GE - Luca Rei
- MIB - Simone Gennai
- NA - Francesco Alessandro Conventi
- PD - Marco Verlato
- PG - Daniele Spiga
- PI - Francesca Lizzi
- ROMA1 - Stefano Giagu



Richieste finanziarie

Aggiornamento e manutenzione della farm: 40 k€ / anno al CNAF

Missioni per *Advanced Hackathon Workshop*:

- 1 k€ / Struttura / anno
- + 4 k€ su Firenze