Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
ICSC

Fourth ML_INFN hackathon
Welcome & introduction

Lucio Anderlini
Istituto Nazionale di Fisica Nucleare
Sezione di Firenze

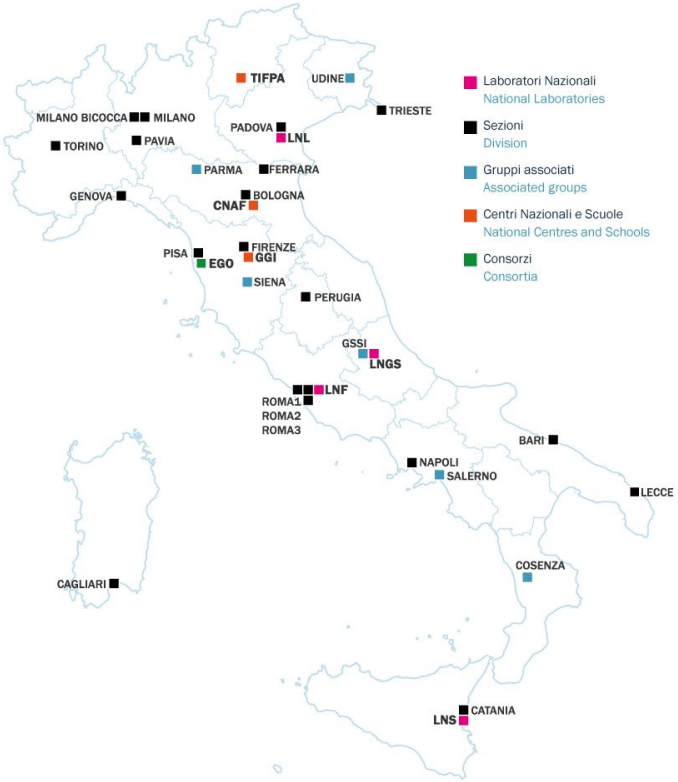
INFN Istituto Nazionale di Fisica Nucleare
SEZIONE DI FIRENZE

m l INFN

INFN CLOUD

ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

INFN Research and structures



216 activities distributed in 33 units (labs, groups and divisions)

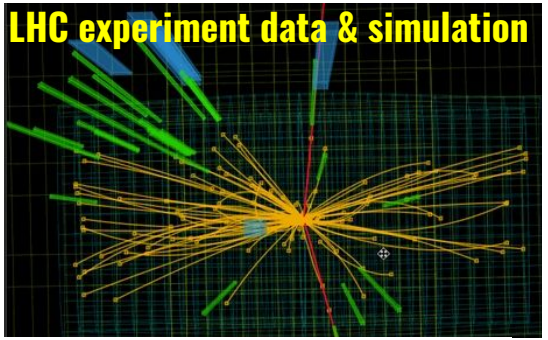
CSN1	**Particle Physics**	17 experiments
CSN2	**Astroparticle Physics**	45 experiments
CSN3	**Nuclear Physics**	23 experiments
CSN4	**Theoretical Physics**	35 initiatives
CSN5	**Technological Research**	96 experiments

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
ICSC

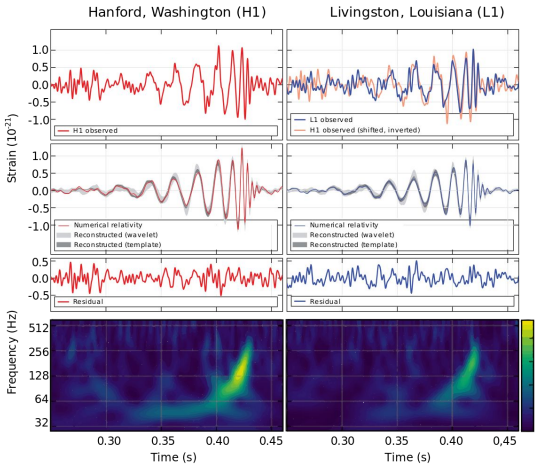
Machine Learning Technologies for INFN

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

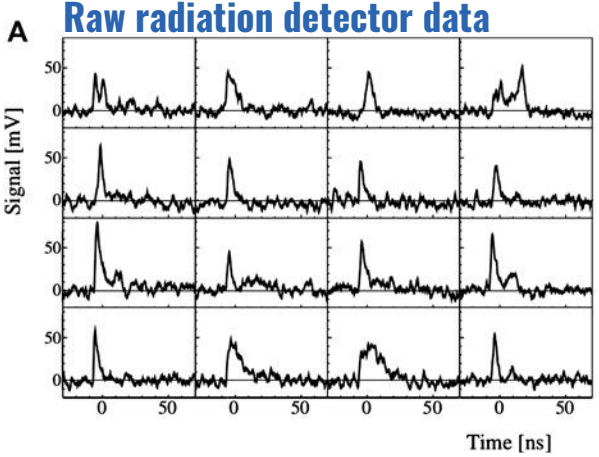
Enthusiasm on the modern data processing technologies!

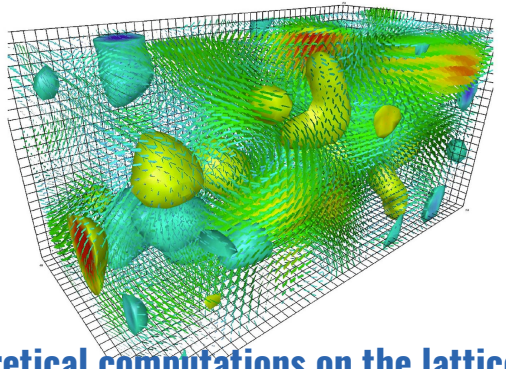


Gravitational wave detection

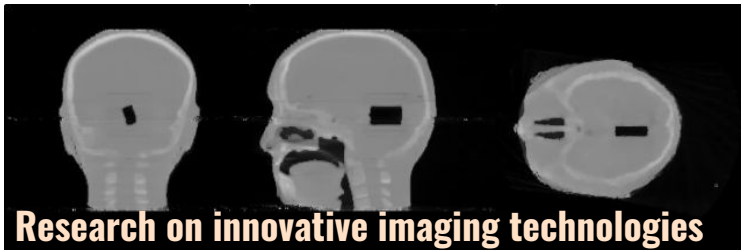


Raw radiation detector data

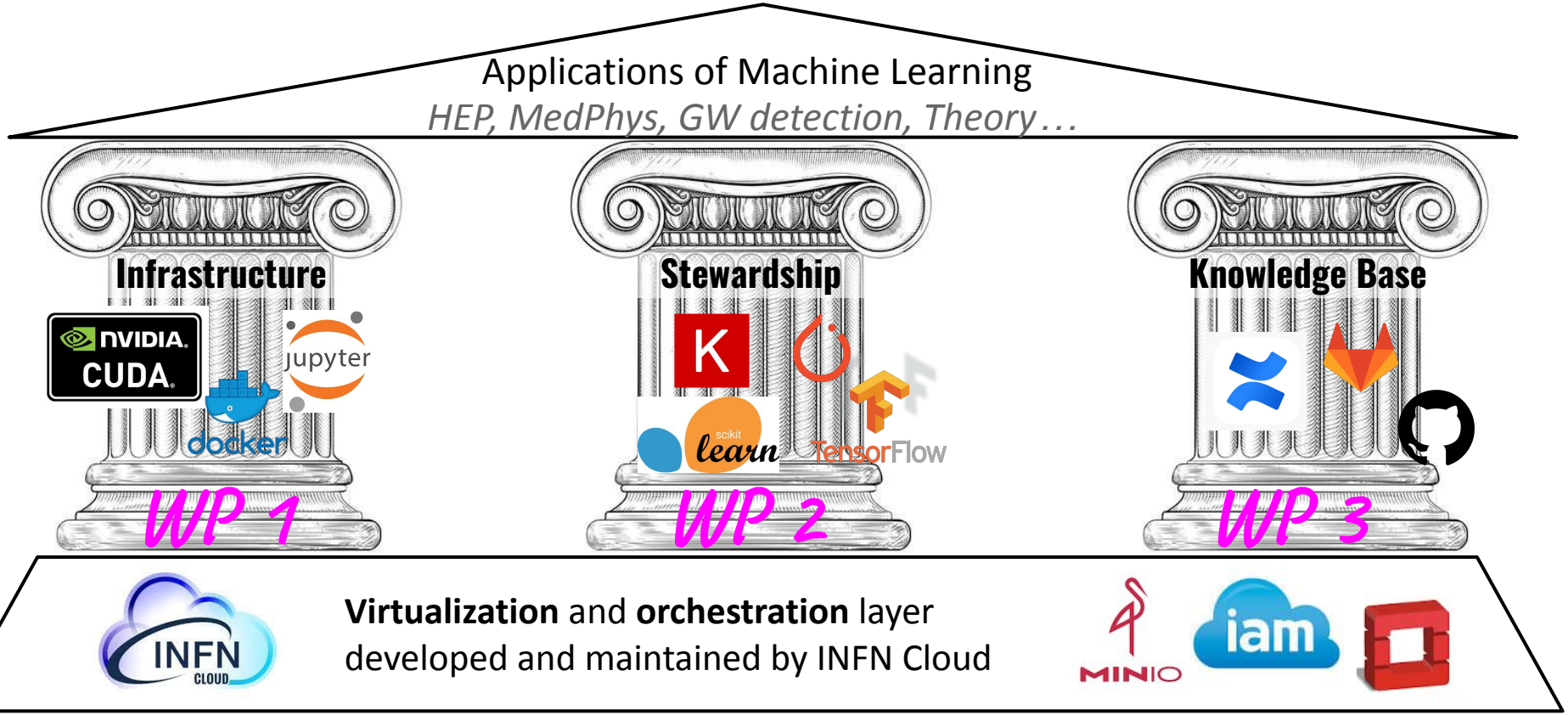




Theoretical computations on the lattice



ML_INFN: The structure of the project



The numbers of ML_INFN

12 INFN **structures** involved in the developments, training activities and hackathons

79 **researchers** devoting a fraction of their time to promote ML techniques for research

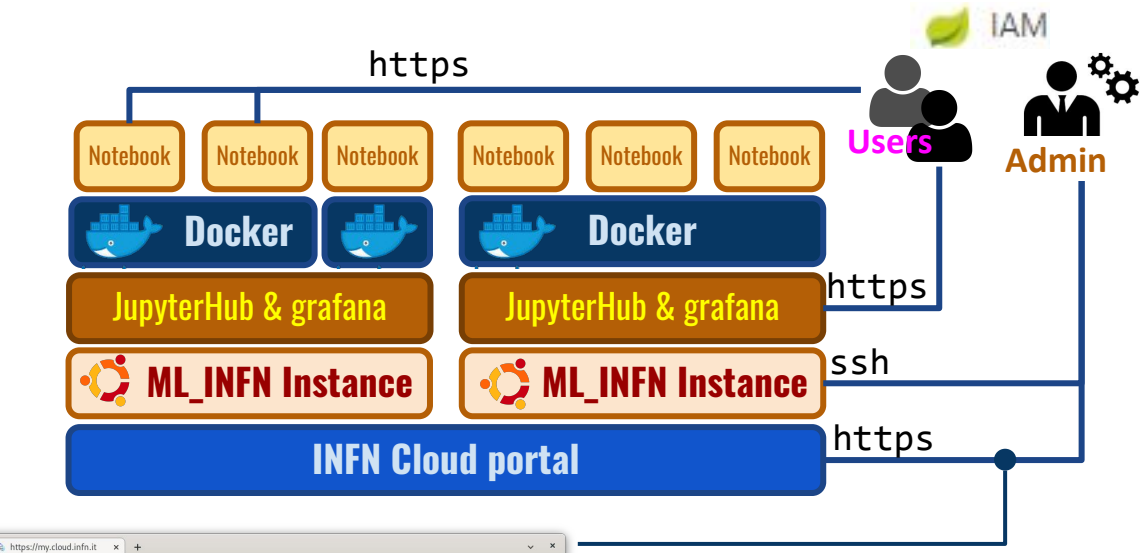
14 professional **GPUs** made available and accessible through the INFN Cloud Interface

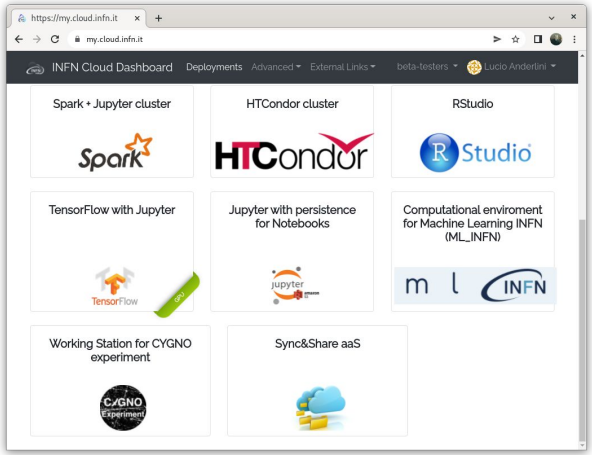
143 **participants** to the **hackathons**, ranging from students to permanent staff members

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
ICSC

INFN Cloud

ML_INFN is built on top of **INFN Cloud**: a data lake-centric, heterogeneous federated Cloud infrastructure spanning multiple sites across Italy, providing an extensible portfolio of solutions tailored to **multidisciplinary scientific communities**.





More detailed in the dedicated talk by Daniele Spiga, today

Federated bare-metal resources

1× SuperMicro + 1× E4 servers:

- 1 TB RAM
- 64-128 CPU cores
- 36 TB local storage (NVMe)
- 8× **Tesla T4** GPUs
- 5× **RTX 5000** GPUs
- 1× **A30** GPU
- 1× **A100** GPU, served as 7 independent MIG slices
- 10 GbE connection to CNAF resources

Federated to CNAF OpenStack and INFN Cloud

Storage solutions

Storage from CERN experiments can be mounted with NFS from the Tier-1 storage

Hypervisors integrated to Ceph to manage persistent virtual volumes accessed from the VM with POSIX

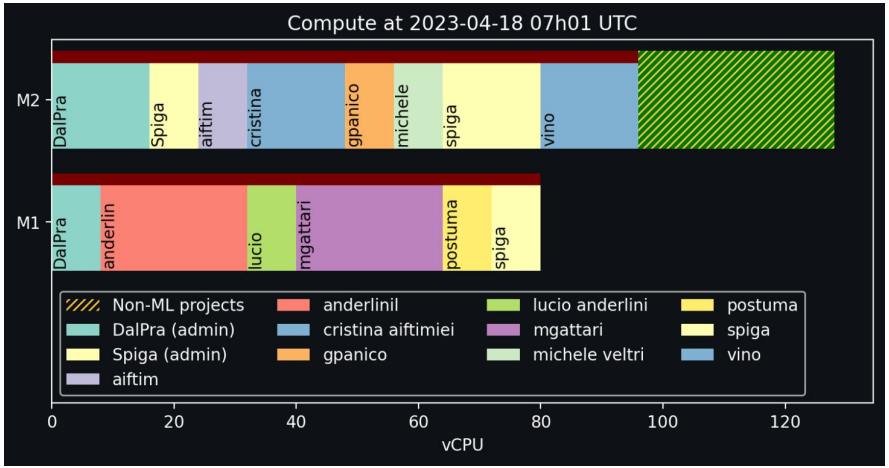
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
ICSC

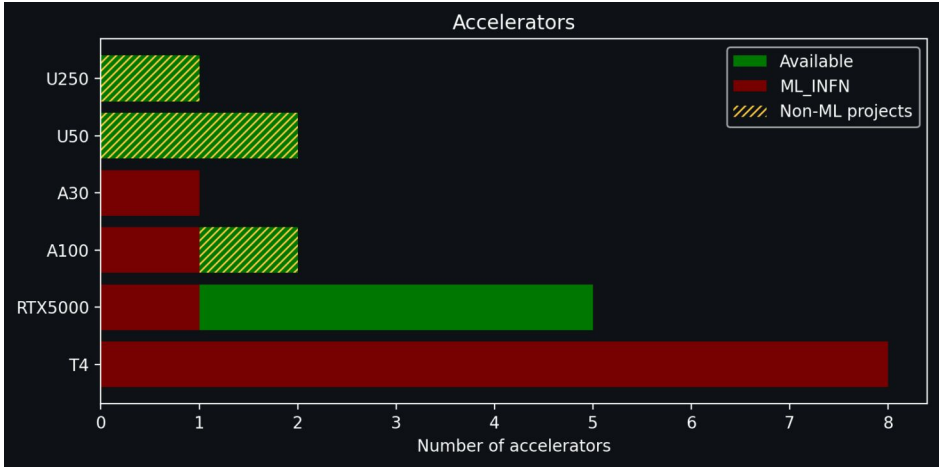
Monitoring and accounting

On top of INFN Cloud services, special tools tuned for ML workflows.

Streamlit dashboard to visualize share of resources at regular snapshots

Grafana dashboard per-VM for utilization.



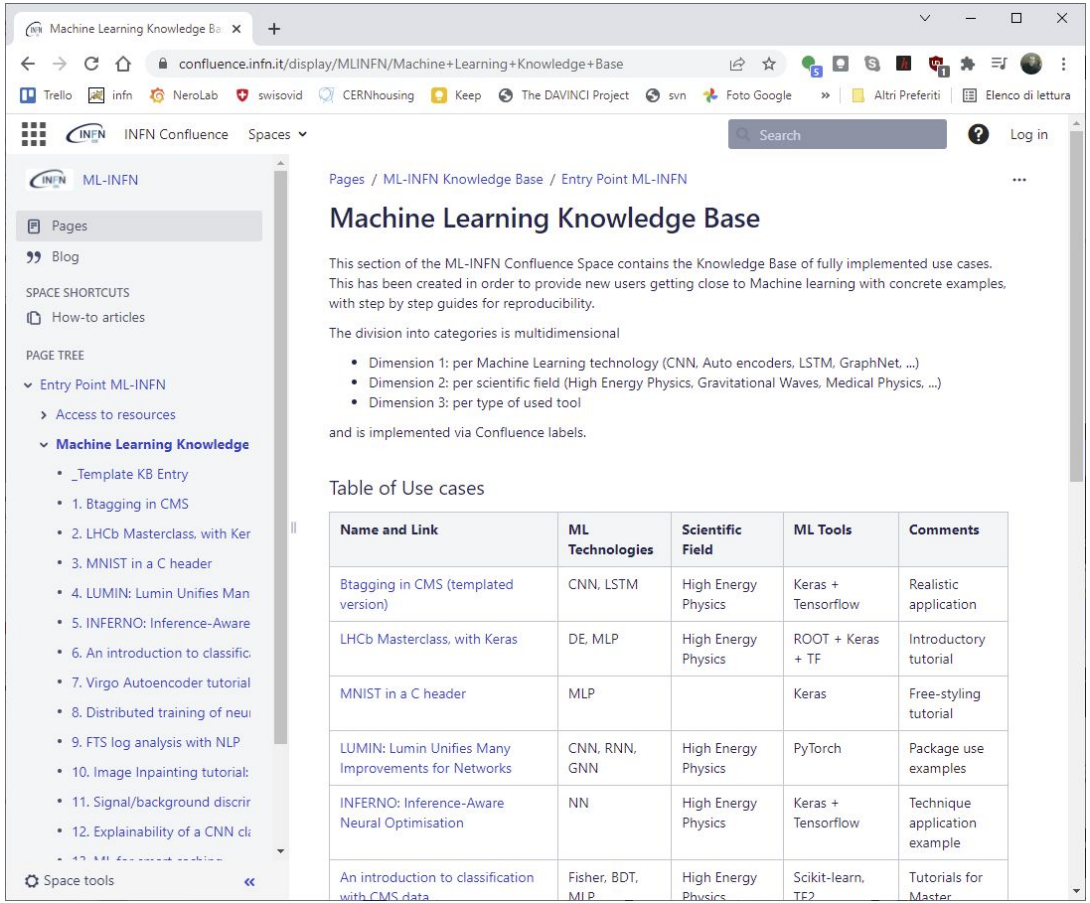


Confluence Knowledge Base

Atlassian Confluence was used to build a **Knowledge Base** reporting several machine-learning use cases, including those discussed at the hackathon.

Each entry includes:

- Runnable **example** as a jupyter notebook or a git repository
- **Contact information** of one or more experts



Machine Learning hackathons: *Base and Advanced level*

To foster the adoption of machine learning tools and techniques in INFN community, we organize events to discuss ML algorithm with the time to look at (and hack) the code.

Starting-level Hackathons

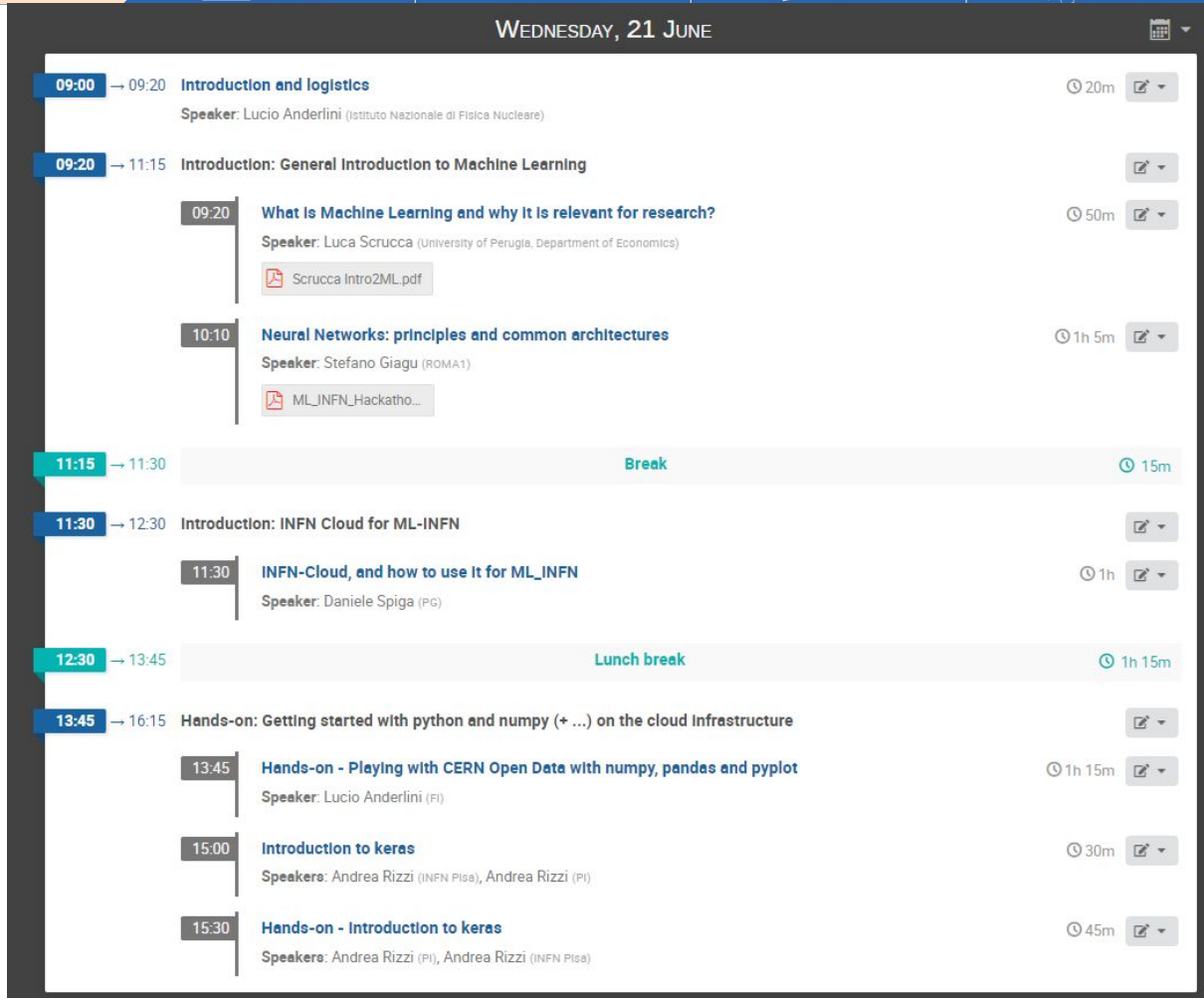
- **online events** with no fee
- up to **60 participants**
- **1 tutor per 5 participants**
- INFN Cloud **CPUs** with shared filesystem

Advanced Hackathons

- **in-person events**
- up to **30 participants**
- **(almost) 1 tutor per participant**
- INFN Cloud **GPUs** with shared filesystem

Day 1

Lectures and introduction to the hands-on



Day 2

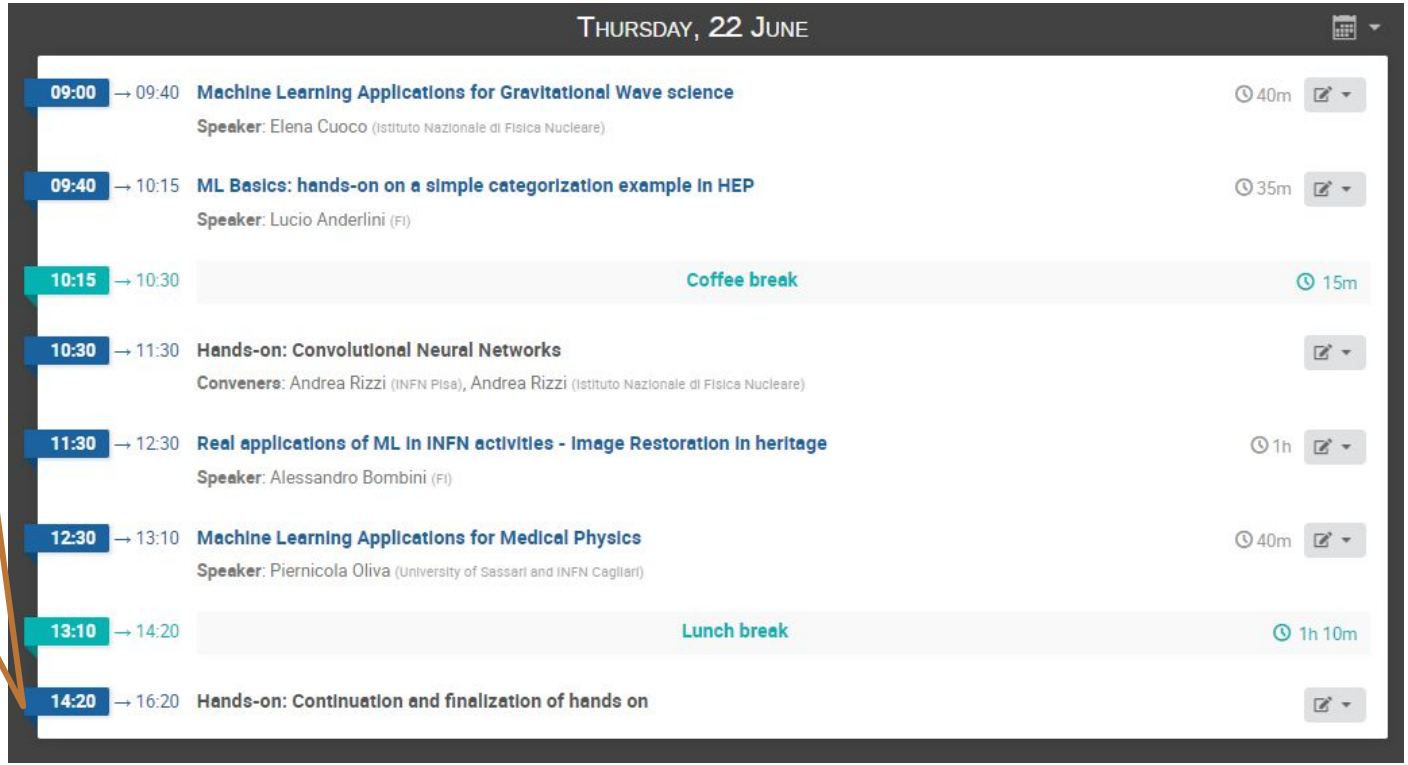
Seminars on real use-cases of machine learning in INFN research and exercises inspired to research use cases

We will suggest exercises during the morning.

During the afternoon you will have the time to go through (some of) them, getting ready for the Friday hackathon.

We will be around for support and discussion.

Ask questions!



THURSDAY, 22 JUNE

09:00 → 09:40	**Machine Learning Applications for Gravitational Wave science** Speaker: Elena Cuoco (Istituto Nazionale di Fisica Nucleare)	40m
09:40 → 10:15	**ML Basics: hands-on on a simple categorization example in HEP** Speaker: Lucio Anderlini (FI)	35m
10:15 → 10:30	Coffee break	15m
10:30 → 11:30	**Hands-on: Convolutional Neural Networks** Conveners: Andrea Rizzi (INFN Pisa), Andrea Rizzi (Istituto Nazionale di Fisica Nucleare)	
11:30 → 12:30	**Real applications of ML in INFN activities - Image Restoration in heritage** Speaker: Alessandro Bombini (FI)	1h
12:30 → 13:10	**Machine Learning Applications for Medical Physics** Speaker: Piernicola Oliva (University of Sassari and INFN Cagliari)	40m
13:10 → 14:20	Lunch break	1h 10m
14:20 → 16:20	**Hands-on: Continuation and finalization of hands on**	

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
ICSC

Day 3

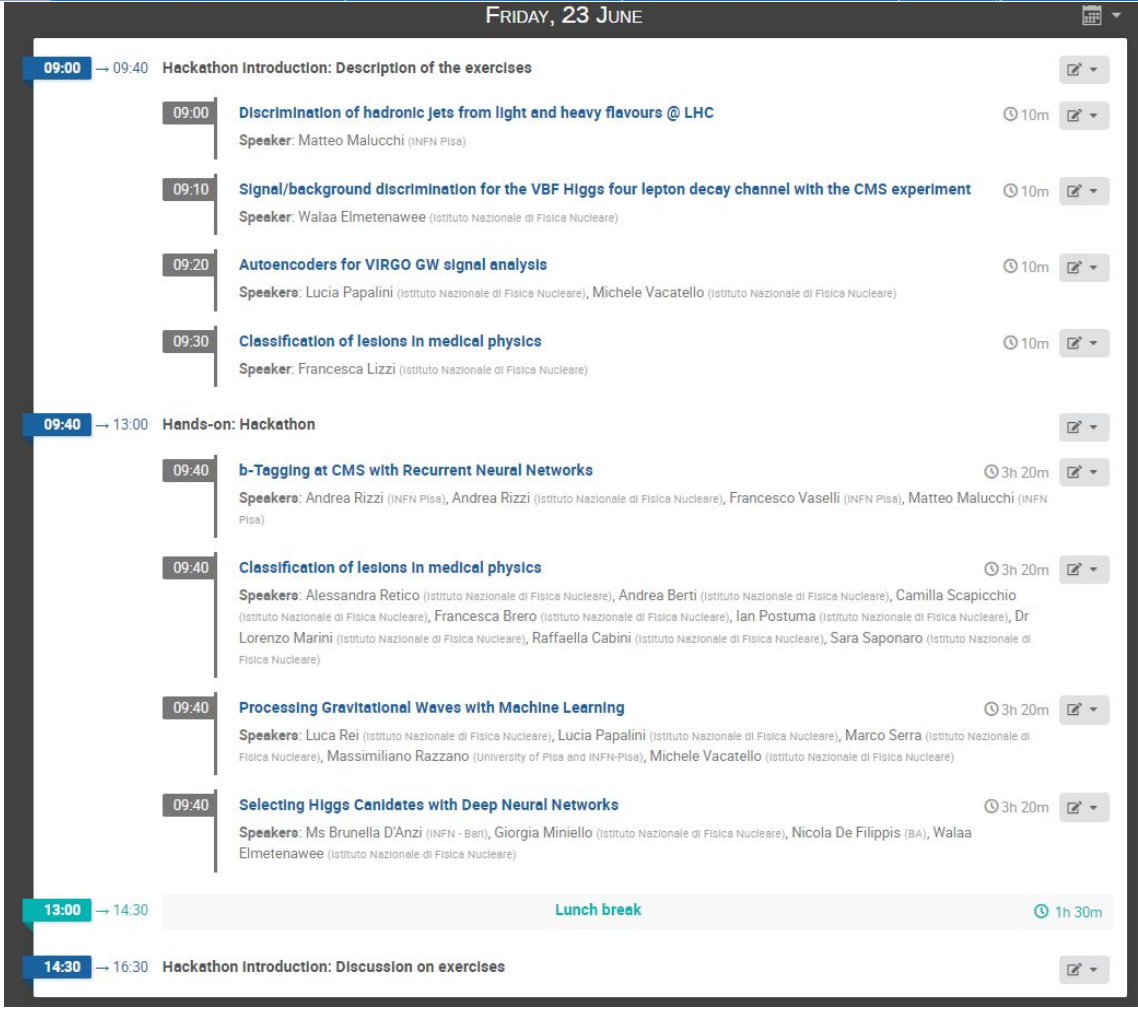
On Friday, the real hackathon.

You will be split in 10 groups.

Each group is assigned to a tutor who will introduce and support your work.

Each group will access a dedicated machine at CNAF.

During the afternoon we will reconvene all together to discuss your solutions and close the event



FRIDAY, 23 JUNE

09:00 → 09:40 Hackathon Introduction: Description of the exercises

09:00 Discrimination of hadronic jets from light and heavy flavours @ LHC — 10m
Speaker: Matteo Malucchi (INFN Pisa)

09:10 Signal/background discrimination for the VBF Higgs four lepton decay channel with the CMS experiment — 10m
Speaker: Walaa Elmetenawee (Istituto Nazionale di Fisica Nucleare)

09:20 Autoencoders for VIRGO GW signal analysis — 10m
Speakers: Lucia Papalini (Istituto Nazionale di Fisica Nucleare), Michele Vacatello (Istituto Nazionale di Fisica Nucleare)

09:30 Classification of lesions in medical physics — 10m
Speaker: Francesca Lizzi (Istituto Nazionale di Fisica Nucleare)

09:40 → 13:00 Hands-on: Hackathon

09:40 b-Tagging at CMS with Recurrent Neural Networks — 3h 20m
Speakers: Andrea Rizzi (INFN Pisa), Andrea Rizzi (Istituto Nazionale di Fisica Nucleare), Francesco Vaselli (INFN Pisa), Matteo Malucchi (INFN Pisa)

09:40 Classification of lesions in medical physics — 3h 20m
Speakers: Alessandra Retico (Istituto Nazionale di Fisica Nucleare), Andrea Berti (Istituto Nazionale di Fisica Nucleare), Camilla Scapicchio (Istituto Nazionale di Fisica Nucleare), Francesca Brero (Istituto Nazionale di Fisica Nucleare), Ian Postuma (Istituto Nazionale di Fisica Nucleare), Dr Lorenzo Marini (Istituto Nazionale di Fisica Nucleare), Raffaella Cabini (Istituto Nazionale di Fisica Nucleare), Sara Saponaro (Istituto Nazionale di Fisica Nucleare)

09:40 Processing Gravitational Waves with Machine Learning — 3h 20m
Speakers: Luca Rei (Istituto Nazionale di Fisica Nucleare), Lucia Papalini (Istituto Nazionale di Fisica Nucleare), Marco Serra (Istituto Nazionale di Fisica Nucleare), Massimiliano Razzano (University of Pisa and INFN-Pisa), Michele Vacatello (Istituto Nazionale di Fisica Nucleare)

09:40 Selecting Higgs Canidates with Deep Neural Networks — 3h 20m
Speakers: Ms Brunella D'Anzi (INFN - Bari), Giorgia Miniello (Istituto Nazionale di Fisica Nucleare), Nicola De Filippis (BA), Walaa Elmetenawee (Istituto Nazionale di Fisica Nucleare)

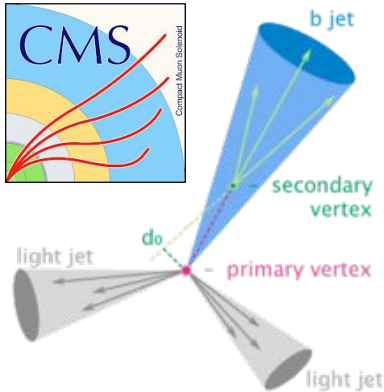
13:00 → 14:30 Lunch break — 1h 30m

14:30 → 16:30 Hackathon Introduction: Discussion on exercises

Hackathon use cases: *10 groups, one tutor per group*

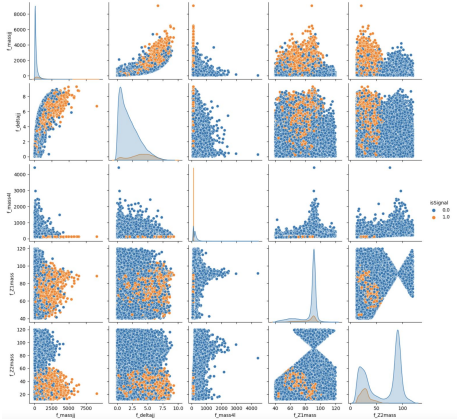
Jet b-tagging at CMS

Recurrent Neural Networks with LSTM



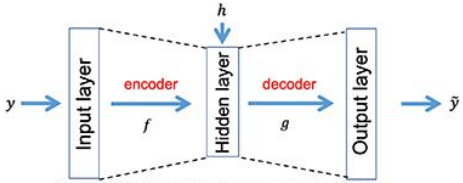
Higgs searches at CMS

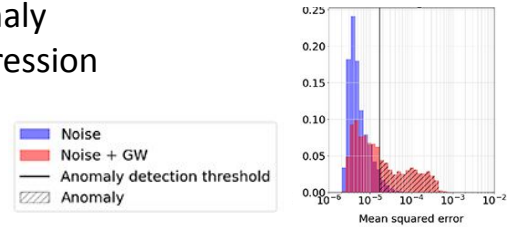
Deep Neural Networks and Advanced Keras



Gravitational Waves with Virgo

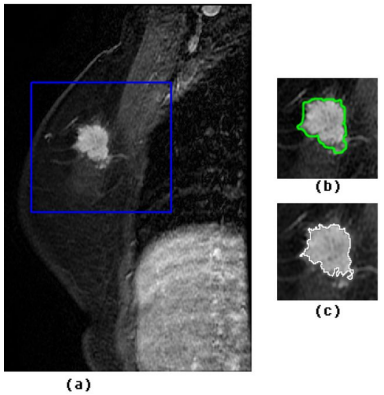
Autoencoders, anomaly detection and compression





Segmentation of CT scans

Convolutional Neural Networks Handling 2D and 3D datasets



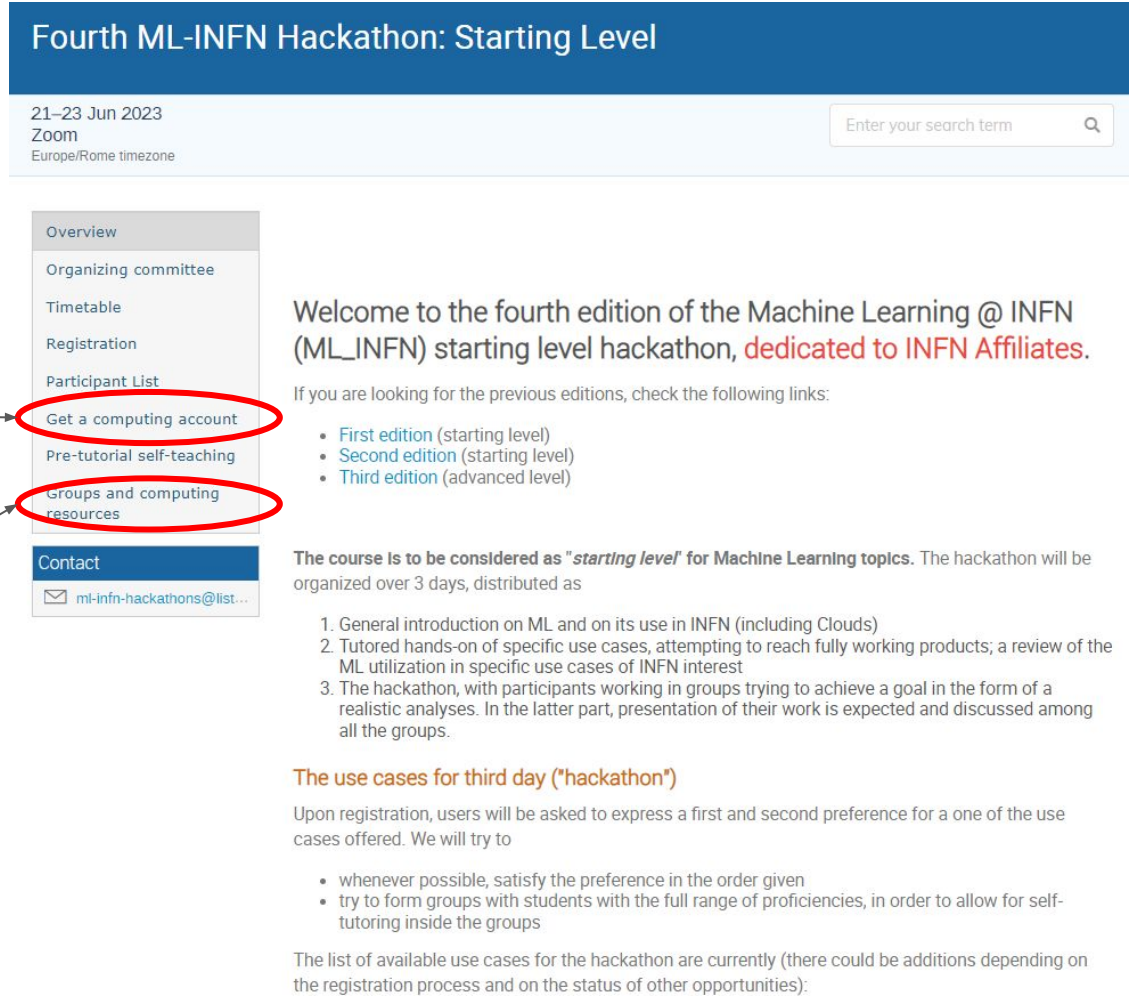
Very important:

Indico page:
agenda.infn.it/event/35607

If you haven't yet,
get your account to access the machines

Check the group (and the machine) assigned to you.

You can use your machine also for the hands-on exercises of day 1 and day 2.



Summary

Welcome to the Fourth ML_INFN Hackathon!

The event is online to lower the costs and enable younger students to get involved.

Let's make an effort to make it equally effective:

1. ask questions (either on zoom or through the mailing list ml-infn-hackathons@lists.infn.it)
2. pick the exercises that you're most curious about and give them a try!
3. have fun!