

# IT ML infrastructure initiative

2nd workshop - October 11th, 2023  
Sofia Vallecorsa, Ricardo Rocha

<https://indico.cern.ch/event/1298990/>

# Motivation

Machine Learning and Deep Learning techniques are investigated for a broad variety of tasks at CERN and in HEP

**ATS:** heavily studying ML/DL for beam control, simulation, engineering, ...

**LHC experiments:** trigger, reconstruction, analysis, simulation...

Different levels of maturity from prototyping stage to production

Need to insure model robustness, reliability, customization, efficient deployment

**IT:** Infrastructure anomaly detection, ...

**Essential to have access to both reliable and scalable software and hardware**

**CERN IT has a role to provide the research community with the needed tools**

# A Machine Learning infrastructure Initiative in IT

ML/DL workloads don't have a monolithic profile.

Depending on task and deployment constraints, computing requirements can be very different.

Map the ML/DL R&D landscape

Lack of resources could hinder algorithm development and the exploration of new research directions

[CERN IT ML Initiative](#), with concrete workshops for infrastructure, use cases,

- Taking into account existing, related efforts at CERN
- Focusing on better understanding current and future needs for ML infrastructure at CERN

Goal:

Evaluate infrastructure and tools used today

Better exposing the current possibilities from CERN IT, both on-premises and public clouds

Proposing synergies and solutions for ML/DL at the medium/long term

Collecting use cases and prototyping using the different available options

# First workshop

<https://indico.cern.ch/event/1253881/>

## →Propose ML initiatives with all business engagement units

To help track requirements from the different communities

To assert the offering in CERN IT is aligned with those requirements

## →Improve GPU sharing between different IT services

Given the scarcity of this type of resource

With the short-term goal of optimizing the overall usage of the existing infrastructure

Note: A short term task, split from the goal of thinking the future ML infrastructure

## →Investigate integration of public cloud resources

Low level (GPUs) as well as higher level services

Consider usage reporting, accounting, etc

## →Ensure ML needs are taken into account in long term resource procurement

Access to public cloud services, procurement of accelerators

## →Improve dataset and model registry and bookkeeping

Both local and external (reusable) models

## CERN IT Machine Learning Infrastructure Workshop

March 10, 2023  
CERN  
Europe/Zurich timezone

Enter your search term

Overview

Timetable

Contribution List

### Timetable

< Fri 10/03

09:00	<b>Welcome and Introduction</b>	Ricardo Rocha et al.		
	31/3-004 - IT Amphitheatre, CERN		09:00 - 09:10	
	<b>SWAN</b>	Diogo Castro et al.		
	31/3-004 - IT Amphitheatre, CERN		09:10 - 09:35	
	<b>Batch and Lxplus</b>	Laurence Field		
	31/3-004 - IT Amphitheatre, CERN		09:35 - 10:00	
10:00	<b>Spark Ecosystem for Machine Learning</b>	Luca Canali		
	31/3-004 - IT Amphitheatre, CERN		10:00 - 10:25	
	<b>Coffee Break</b>			
	31/3-004 - IT Amphitheatre, CERN		10:25 - 10:40	
	<b>Kubeflow - ml.cern.ch</b>	Dejan Golubovic et al.		
	31/3-004 - IT Amphitheatre, CERN		10:40 - 11:05	
11:00	<b>Public Cloud</b>	Dr Sofia Vallecorsa		
	31/3-004 - IT Amphitheatre, CERN		11:05 - 11:30	
	<b>Application: ADMON</b>	Nikolay Tsvetkov		
	31/3-004 - IT Amphitheatre, CERN		11:30 - 11:37	
	<b>Application: UNOSAT</b>	Edoardo Nemmi		
	31/3-004 - IT Amphitheatre, CERN		11:37 - 11:44	
	<b>Application: Cloud Anomaly Detection</b>	Domenico Giordano		
	31/3-004 - IT Amphitheatre, CERN		11:44 - 11:51	
	<b>Application: Network traffic estimation</b>	Edoardo Martelli et al.		
	31/3-004 - IT Amphitheatre, CERN		11:51 - 11:58	
12:00	<b>Application: Foundation Models</b>	Renato Paulo Da Costa Cardoso		
	31/3-004 - IT Amphitheatre, CERN		11:58 - 12:05	
	<b>Application: Digital twins</b>	Alexander Zuehbauer et al.		
	31/3-004 - IT Amphitheatre, CERN		12:05 - 12:12	
	<b>Application: REANA</b>	Tibor Simko		
	31/3-004 - IT Amphitheatre, CERN		12:12 - 12:19	
	<b>Discussion and Next Steps</b>			
	31/3-004 - IT Amphitheatre, CERN		12:20 - 12:45	

# Today

## 2nd CERN IT ML Infrastructure Workshop

Focusing on representative use cases and CERN wide requirements

### 2nd CERN IT Machine Learning Infrastructure Workshop

Wednesday 11 Oct 2023, 09:00 → 13:00 Europe/Zurich  
40/S2-A01 - Salle Anderson (CERN)

**Description** The High Energy Physics community has started introducing Deep Learning techniques for improving different aspects in the experiments and accelerators life cycle and data processing steps.

The availability of reliable, user-friendly and scalable resources, including full software and hardware stacks, is critical to fully support activities in the ML/DL domain. In this context, the CERN IT department has launched an initiative to gather information about the status of its ML/DL dedicated infrastructure, the state-of-the-art and the needs of the HEP ML/DL community at CERN.

A first workshop focused on the current status of ML/DL activities and infrastructure solutions in CERN IT. This second workshop tries to collect information about on-going and planned AI-research within the different departments and experiments at CERN.

**Videoconference** 2nd CERN IT Machine Learning Infrastructure Workshop

09:00	→ 09:10	<b>Introduction</b>	🔊 10m	🗒️
Speakers: Ricardo Rocha (CERN), Dr Sofia Vallecorsa (CERN)				
09:10	→ 09:30	<b>ATLAS</b>	🔊 20m	🗒️
Speaker: Daniel Thomas Murnane (Lawrence Berkeley National Lab. (US))				
09:30	→ 09:50	<b>CMS</b>	🔊 20m	🗒️
Speaker: Davide Valsecchi (ETH Zurich (CH))				
09:50	→ 10:10	<b>ALICE</b>	🔊 20m	🗒️
Speaker: Fabio Catalano (CERN)				
10:10	→ 10:30	<b>LHCb</b>	🔊 20m	🗒️
Speaker: Simon Akar (University of Cincinnati (US))				
10:30	→ 10:50	<b>ATS</b>	🔊 20m	🗒️
Speaker: Florian Rehm (CERN / RWTH Aachen University)				
10:50	→ 11:10	<b>Break</b>	🔊 20m	
11:10	→ 12:10	<b>Discussion</b>	🔊 1h	🗒️

# Next

We are organising a set of talks/seminar inviting experts from other communities or industries to collect best practices in different domains

Questions?