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

UNLOCKING THE COMPUTE CONTINUUM

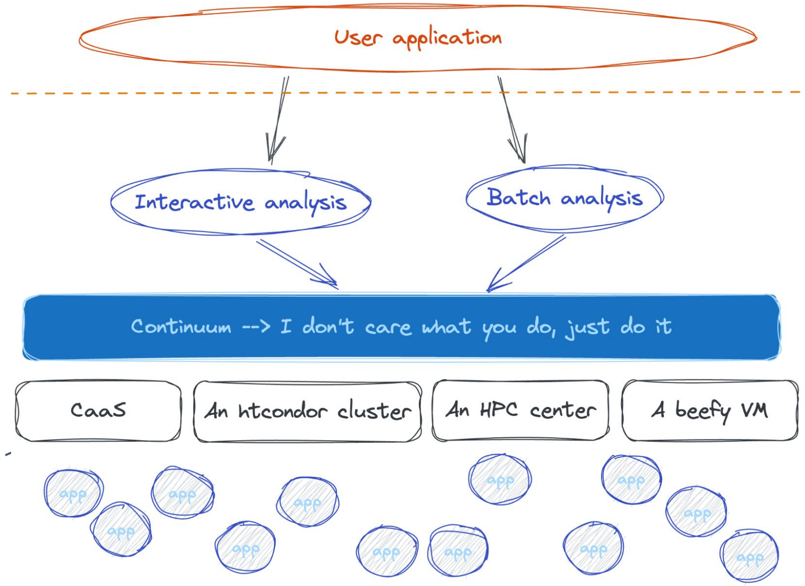
SCALING OUT FROM CLOUD TO H{P,T}C RESOURCES

Diego Ciangottini -INFN
Tommaso Boccali - INFN
Lucio Anderlini - INFN
Andrej Filipcic - JSI
Teo Prica - IZUM
Tommaso Tedeschi - INFN
Daniele Spiga - INFN
Massimo Sgaravatto - INFN
Ahmed Shiraz Memon - JUELICH
Andrea Manzi - EGI
Antonino Troja - INFN
Giulio Bianchini - INFN
Federica Fanzago - INFN

CONTINUUM

Heterogeneous use cases

- **Coffee-break time long analyses at LHC**
  - cloud + scale out in high rate frameworks
- **GPU accelerated HEP simulations and triggers**
  - develop on cloud, deploy at scale (on grid?)
- **ML training and inference of particle tagging**
  - Cloud based data-science frameworks hungry for HPC hosted GPUs
- **Well known grid-friendly workflows**



Our continuum idea is here in the middle

Heterogeneous resources managed by heterogeneous systems

HPC → mainly SLURM
HTC → any grid batch system (HTCondor, ARC, ...)
Cloud → Kubernetes

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

Contribution List

18 / 512 ML

568. Welcome
Agnieszka Dziurda (Polish Academy of Sciences (PL)), Tomasz Szumlak (AGH University of Krakow (PL))
21/10/2024, 09:00
Talk | Plenary session

114. CMS FlashSim: end-to-end simulation with ML
Andrea Rizzi
21/10/2024, 11:30
Plenary | Talk | Plenary session
Detailed event simulation at the LHC is taking a large fraction of computing budget. CMS developed an end-to-end ML based simulation that can speed up the time for production of analysis samples of several orders of magnitude with a

380. Fair Universe HiggsML Uncertainty Challenge
Yulei Zhang (Shanghai Jiao Tong University (CN) & APC-Paris (FR))
21/10/2024, 13:48
Track 5 - Simulation and ... | Talk | Parallel (Track 5)
The Fair Universe project is organising the HiggsML Uncertainty Challenge, which will/has run from June to October 2024.

60. S3 Compatibility: Enabling Seamless Integration with EOS, CERN's Large-Scale Disk Storage System (MON 30)
Andreas Joachim Peters (CERN), Elvin Alin Sindrilaru (CERN), Luca Mascetti (CERN)
21/10/2024, 15:18
Track 1 - Data and Meta... | Poster | Poster session
Amazon S3 is a leading object storage service known for its scalability, data reliability, security and performance. It is used as a storage solution for data lakes, websites, mobile applications, backup, archiving and more. With its

20. Streamlining ATLAS Monte-Carlo Generator Validation with PAVER (MON 33)
Mustafa Andre Schmidt (Bergische Universitaet Wuppertal (DE))
21/10/2024, 15:18
Track 5 - Simulation and ... | Poster | Poster session
In ATLAS and other high-energy physics experiments, the integrity of Monte-Carlo (MC) simulations is crucial for reliable physics analysis. The continuous evolution of MC generators necessitates regular validation to ensure the accuracy of

279. End-to-end ML-based reconstruction
Dolores Garcia (CERN)
22/10/2024, 14:06
Track 3 - Offline Comput... | Talk | Parallel (Track 3)
We present an ML-based end-to-end algorithm for adaptive reconstruction in different FCC detectors. The algorithm takes detector hits from different subdetectors as input and reconstructs higher-level objects. For this, it exploits a

293. ML-based classification of photons and neutral mesons for direct photon measurement in ALICE
Abhishek Nath (Heidelberg University (DE))
22/10/2024, 14:24
Track 5 - Simulation and ... | Talk | Parallel (Track 5)
Direct photons are unique probes to study and characterize the quark-gluon plasma (QGP) as they leave the collision medium mostly unscathed. Measurements at top Large Hadron Collider (LHC) energies at low pT reveal a very small

176. ML-based classification in an open-source framework for the ALICE heavy-flavour analysis (TUE 27)
Maria Teresa Camerlingo (Universita e INFN, Bari (IT))
22/10/2024, 15:18
Track 5 - Simulation and ... | Poster | Poster session
The ALICE Collaboration aims to precisely measure heavy-flavour (HF) hadron production in high-energy proton-proton and heavy-ion collisions since it can provide valuable tests of perturbative quantum chromodynamics models and

119. Simulating the CMS High Granularity Calorimeter with ML
Kevin Pedro (Fermi National Accelerator Lab. (US))
22/10/2024, 16:33
Track 5 - Simulation and ... | Talk | Parallel (Track 5)
Detector simulation is a key component of physics analysis and related activities in CMS. In the upcoming High Luminosity LHC era, simulation will be required to use a smaller fraction of computing in order to satisfy resource

356. Benchmark Studies of ML Inference with TMVA SOFIE
Lorenzo Moneta (CERN)
23/10/2024, 13:30
Track 5 - Simulation and ... | Talk | Parallel (Track 5)
Within the ROOT/TMVA project, we have developed a tool called SOFIE, that takes externally trained deep learning models in ONNX format or Keras and PyTorch native formats and generates C++ code that can be easily included and

271. ML-Assisted Charged Particle Tracking at GlueX
Torri Jeske
23/10/2024, 14:24
Track 2 - Online and real-... | Talk | Parallel (Track 2)
Tracking charged particles resulting from collisions in the presence of strong magnetic field is an important and challenging problem. Reconstructing the tracks from the hits created by those generated particles on the detector layers

323. A Streamlined Neural Model for Real-Time Analysis at the First Level of the LHCb Trigger (WED 06)
Jiahui Zhuo (Univ. of Valencia and CSIC (ES)), Volodymyr Svintozelskyi (Univ. of Valencia and CSIC (ES))
23/10/2024, 15:18
Track 2 - Online and real-... | Poster | Poster session
One of the most significant challenges in tracking reconstruction is the reduction of "ghost tracks," which are composed of false hit combinations in the detectors. When tracking reconstruction is performed in real-time at 30 MHz, it

110. ML-based Adaptive Prefetching and Data Placement for US HEP systems
Dr Byrav Ramamurthy (University of Nebraska-Lincoln)
23/10/2024, 16:51
Track 1 - Data and Meta... | Talk | Parallel (Track 1)
Although caching-based efforts [1] have been in place in the LHC infrastructure in the US, we show that integrating intelligent prefetching and targeted dataset placement into the underlying caching strategy can improve job efficiency

514. Efficient ML-Assisted Particle Track Reconstruction Designs
Nadezhda Dobreva
24/10/2024, 13:48
Track 3 - Offline Comput... | Talk | Parallel (Track 3)
Track reconstruction, a.k.a., tracking, is a crucial part of High Energy Physics experiments. Traditional methods for the task, relying on Kalman Filters, scale poorly with detector occupancy. In the context of the upcoming High Luminosity-

338. Efficiency, Reproducibility, and Portability in HEP Machine Learning Training - ML Training Facility at Vanderbilt University
Andrew Malone Melo (Vanderbilt University (US))
24/10/2024, 14:42
Track 9 - Analysis faciliti... | Talk | Parallel (Track 9)
The success and adoption of machine learning (ML) approaches to solving HEP problems has been widespread and fast. As useful a tool as ML has been to the field, the growing number of applications, larger datasets, and increasing

461. LbMCSubmit: Streamlined production and submission of LHCb MC requests (THU 20)
Emir Muhammad (University of Warwick (GB))
24/10/2024, 15:18
Track 4 - Distributed Co... | Poster | Poster session
The LHCb experiment requires a wide variety of Monte Carlo simulated samples to support its physics programme. LHCb's centralised production system operates on the DIRAC backend of the WLCG; users interact with it via the DIRAC

183. QiboML: a full-stack quantum machine learning framework
Matteo Robbiati (Università degli Studi e INFN Milano (IT))
24/10/2024, 16:51
Track 5 - Simulation and ... | Talk | Parallel (Track 5)
We present QiboML, an open-source software library for Quantum Machine Learning (QML) integrated with the Qibo quantum computing middleware framework.

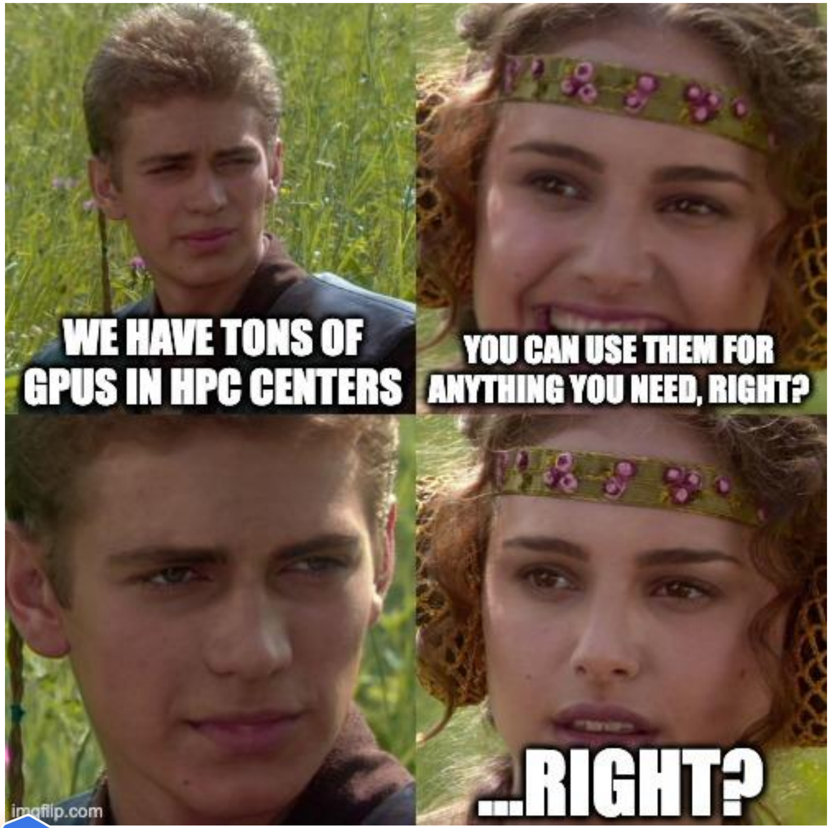
358. Use of topological correlations in ML-based conditions for the CMS Level-1 Global Trigger upgrade for the HL-LHC
Gabriele Bortolato (Universita e INFN, Padova (IT))
24/10/2024, 17:27
Track 2 - Online and real-... | Talk | Parallel (Track 2)
The High-Luminosity LHC upgrade will have a new trigger system that utilizes detailed information from the calorimeter, muon and track finder subsystems at the bunch crossing rate, which enables the final stage of the Level-1 Trigger, the

Are ML and GPUs meaningful use cases?

You can repeat the exercise searching for "GPU"...

Where are my resources?

I need my accelerators!!!



We could/should(?) use HPC supercomputers…

Technically we can already
BUT how about making **a painless experience out of it?**

Can we ignore that most of new data science tools have now a "lingua franca"?
Kubernetes API support is arguably a de-facto standard

Where it all began

A long experience at INFN with site extensions

Two opportunities:

- **ICSC**
- **interTwin EU project**
    - visit EGI booth to know more!

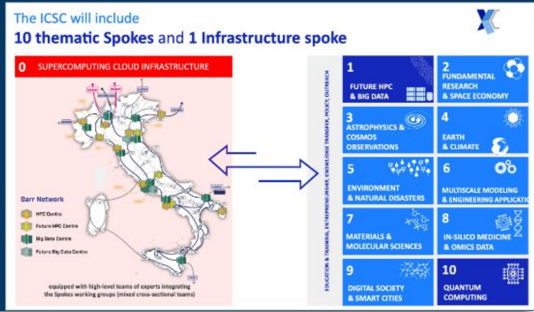
Not only HEP, a good solution is an **adaptable solution**

Everyone should be able to come and plug its own integration logic.

Several heterogeneous science communities: different domains, spanning O(10) to O(1000) people

On a geo-distributed heterogeneous set of computing centers: different resource management (batch, cloud), different size and specialized hardware

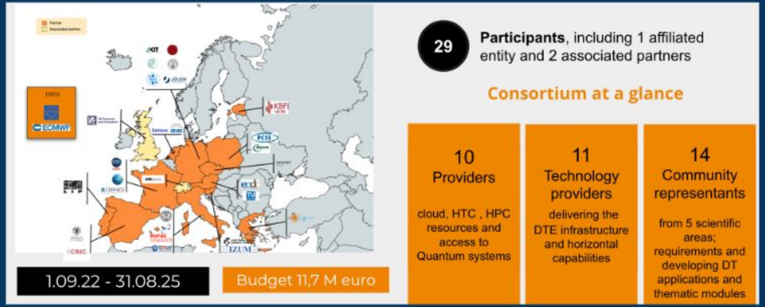
We strive to get an efficient interface to match the evolution of analysis frameworks toward cloud-native tools with a seamless resource brokering.



Very similar set of requirements with an even stronger need for cloud native framework integration with HPC centers (kind of a cloud-prohibitive environment)

A Digital Twin Engine should be able to provide a platform capable of distributing DT applications to remote sites, maintaining an interoperable (cloud) interface.

interTwin

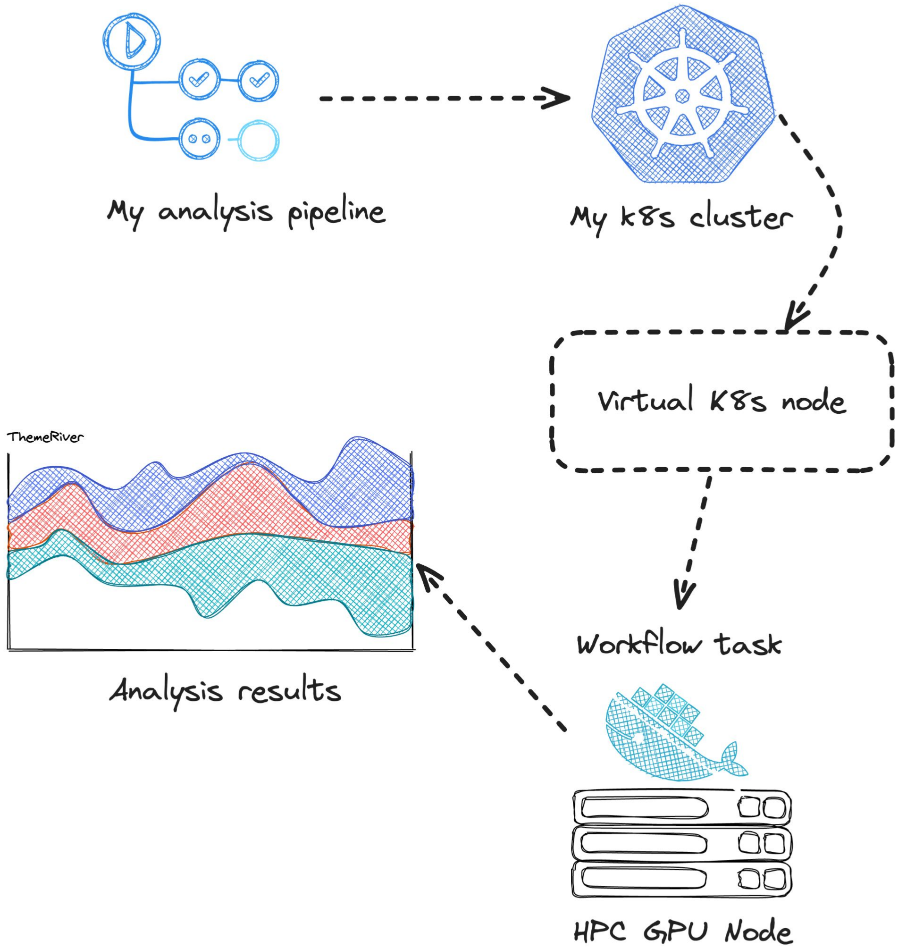


Our strategy

What if we extend the container orchestration layer (K8s) to support "offloading"* under the hood?
(With little to no knowledge required from the end user perspective)

Exposing the **very SAME experience of running a pod** on the cloud resources -> never touching the native API layer by K8s

N.B. Using Kubernetes as the workhorse for the "offloading", NOT as the main user interface though



*"Offloading" refers to the process of delegating the execution of a container to a remote resource instead of a physical node in your cluster.

Enabling tech: virtual nodes!

CNCF Sandbox project

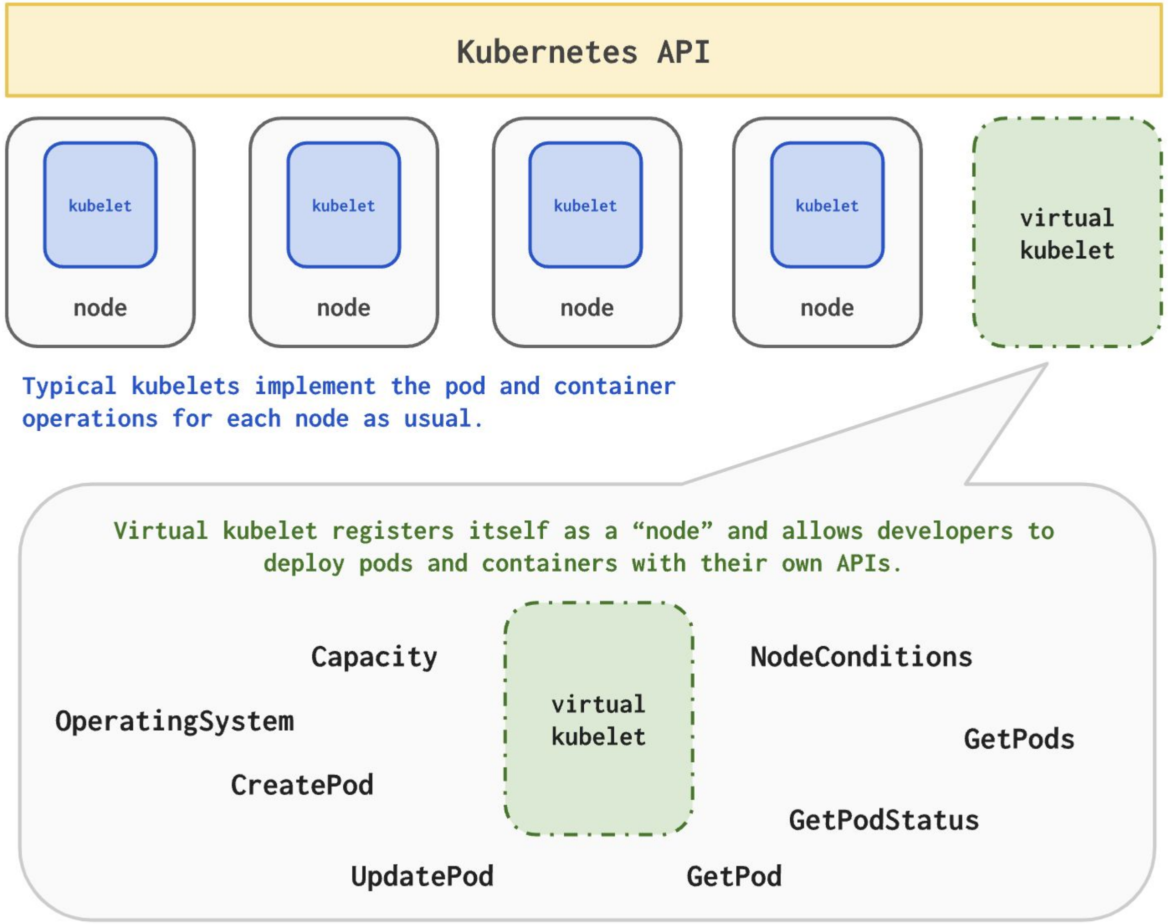
Virtual Kubelet

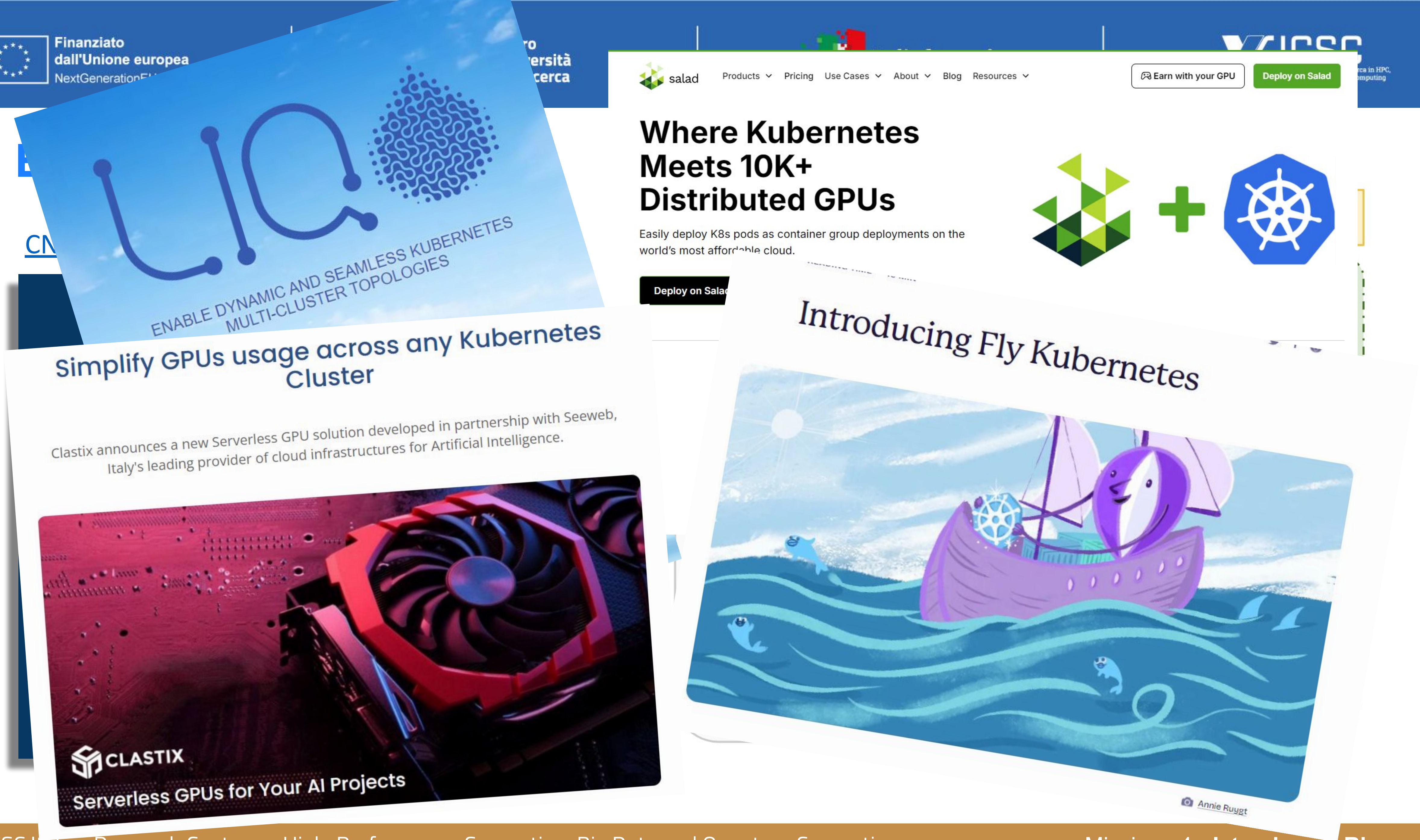
Create a Kubernetes cluster node, but virtual...

Translate a pod request in a custom container execution (e.g. remotely run apptainer/podman)

1. KEEP KUBERNETES API
2. EXPLOIT EXTERNAL/REMOTE RESOURCES

Quite a good fit 😎




liqo
ENABLE DYNAMIC AND SEAMLESS KUBERNETES
MULTI-CLUSTER TOPOLOGIES
Simplify GPUs usage across any Kubernetes Cluster
Clastix announces a new Serverless GPU solution developed in partnership with Seeweb,
Italy's leading provider of cloud infrastructures for Artificial Intelligence.
CLASTIX
Serverless GPUs for Your AI Projects
salad
Products
Pricing
Use Cases
About
Blog
Resources
Earn with your GPU
Deploy on Salad
Where Kubernetes Meets 10K+ Distributed GPUs
Easily deploy K8s pods as container group deployments on the world's most affordable cloud.
Deploy on Salad
Introducing Fly Kubernetes
Annie Ruygt

We need a bit more though

With current Virtual Kubelet interface:

- **Deep knowledge of Kubernetes internals is needed**
  - When developing a different flavors of virtual kubelet "logic"
- **We want providers to be in control**
  - abstracting the layer of resource provisioning from the payload brokering itself

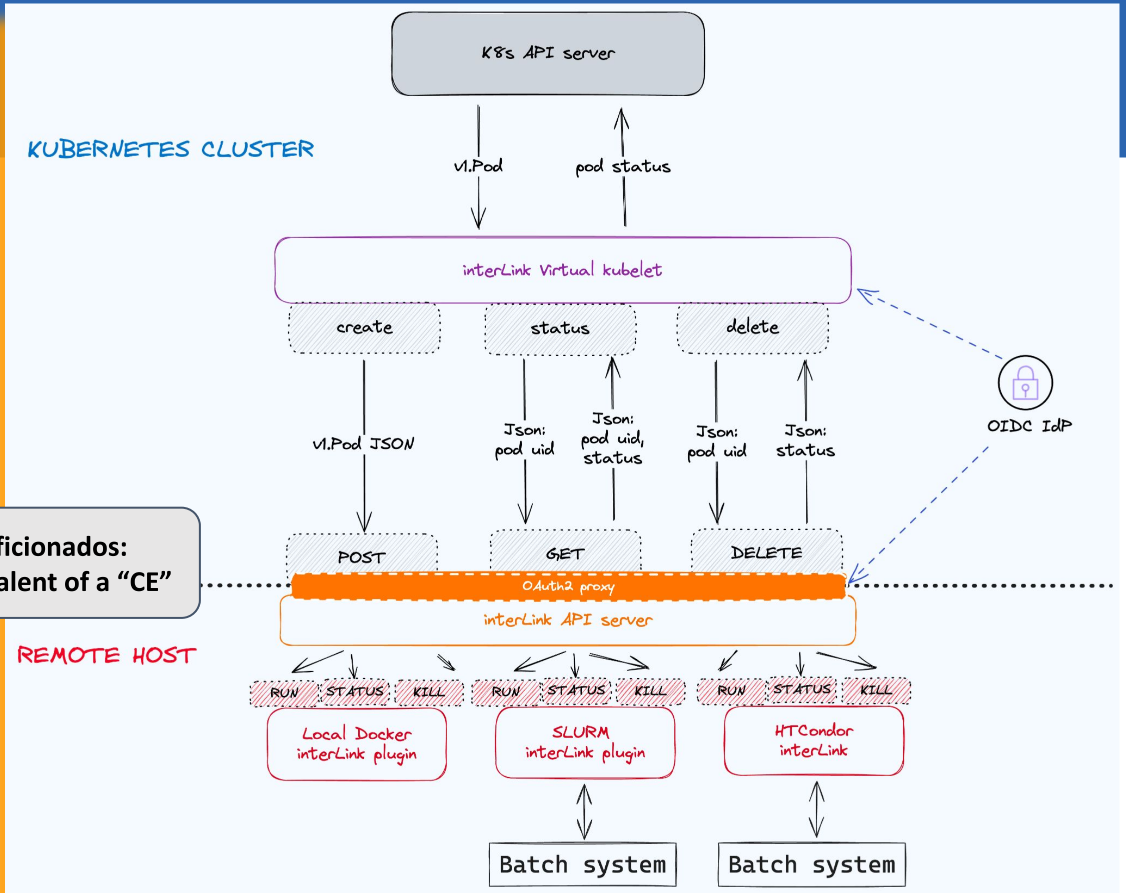
Requiring Kubernetes black belt for any provider administrator of ours is a NO GO

The result is the development of <u>a modular interface to create custom Virtual Kubelets:</u>

interLink

We set a contract

For HTCondor aficionados:
a Kubernetes equivalent of a "CE"



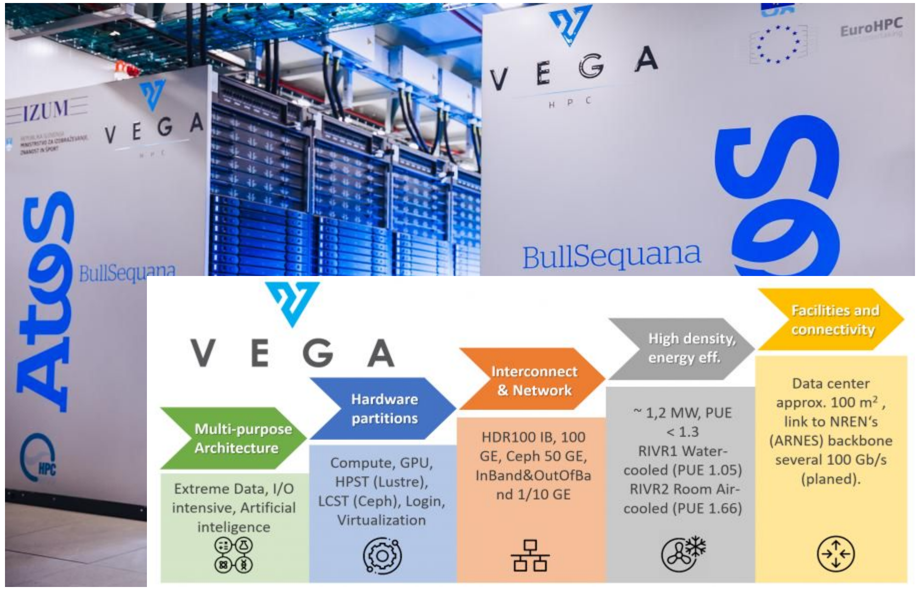
As a community effort we maintain the complexities

VK-interLink communication, AuthN/Z, VK state machine, caching etc..

Providers focus on creating value via additional integrations

All of that without a PhD in k8s, leveraging simple SDKs based on interLink OpenAPI spec

Our first case studies



HPC Vega is the first EuroHPC JU supercomputer hosted at the Institute of Information Science in Maribor, in Slovenia.

The Jülich Supercomputing Centre operates one of the most powerful supercomputers in Europe, JUWELS, and JUNIQ the first European infrastructure for quantum computing. UNICORE offers seamless access to the Supercomputers.

First volunteer HPC provider, enabling super early prototyping

First volunteer for an external plugin based on UNICORE

In both cases the payloads will be submitted to SLURM by the interLink plugin hosted on a VM on the edge of the HPC

Did it work?

First use case driven by **3D GAN algorithm** developed at **CERN** and integrated with **MLFlow tracking**.

Training task offloaded to GPUs available at the HPCs as **"normal" K8s pods**

Pod annotations are a powerful tool to send dedicated requests to the remote container manager!

itwinai-cern-use-case-training.yaml U ×

containers > itwinai-cern-use-case-training.yaml

```yaml
apiVersion: v1
kind: Pod
metadata:
  name: itwinai-cern-use-case-training-run0
  annotations:
    slurm-job.vk.io/flags: "--gres=gpu:4 \
                            --job-name=itwinai \
                            --output=.local/interlink/jobs/itwinai/job.out \
                            --account=intertwin  \
                            --mail-type=ALL \
                            --time=20:00:00"
spec:
  restartPolicy: Never
  containers:
  - image: /p/project/intertwin/zoechbauer1/T6.5-AI-and-ML/use-cases/3dgan/containers/cern.sif
    command: ["/bin/bash"]
    args: ["-c", "\"cd /workspace/T6.5-AI-and-ML/use-cases/3dgan && python train.py -p pipeline.yaml\""]
    imagePullPolicy: Always
    name: ai
  dnsPolicy: ClusterFirst
  nodeSelector:
    kubernetes.io/hostname: jul-vk
  tolerations:
  - key: virtual-node.interlink/no-schedule
    operator: Exists
```

```
[memon1@hdfml101 ~]$ squeue --me
         JOBID PARTITION     NAME     USER ST       TIME  NODES NODELIST(REASON)
         60936     batch Prototyp   memon1 CF       1:28      1 hdfmlc03
```

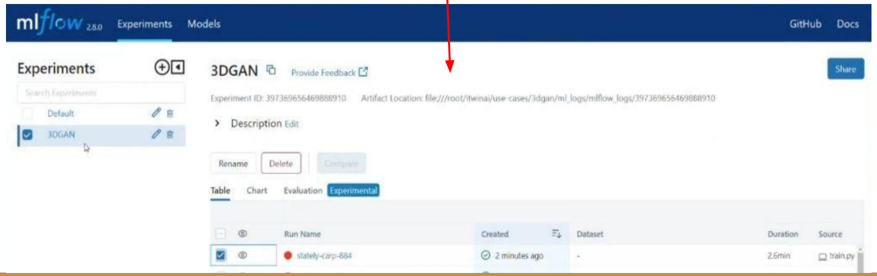
```
[ciangottinid@intertwin ~]$ squeue --me -i5
Tue Sep 26 09:41:12 2023
             JOBID PARTITION     NAME     USER ST       TIME  NODES NODELIST(REASON)
           4086045       cpu test-pod ciangott  R 1-00:41:14      1 cn0130
```

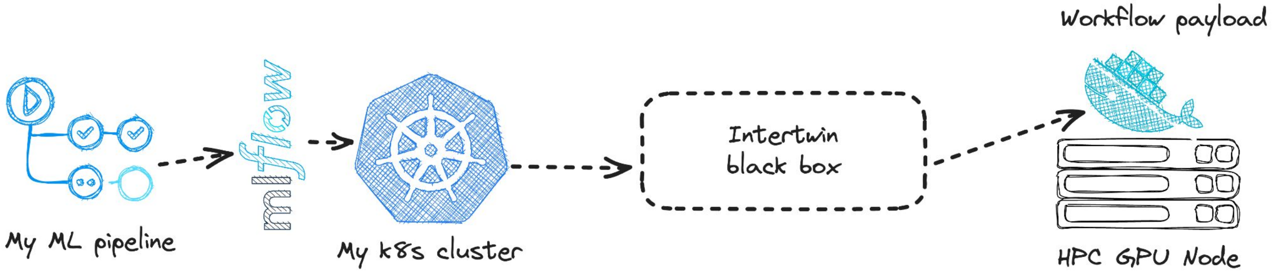
JÜLICH Forschungszentrum

VEGA

```
##############################################
ting execution of 'Lightning3DGANPredictor'... #
##############################################
ing in : /tmp/data/exp_data/*/*.h5
1 files.

Predicting: |          | 0/? [00:00<?, ?it/s]
Predicting:   0%|          | 0/2 [00:00<?, ?it/s]
Predicting DataLoader 0:   0%|          | 0/2 [00:00<?, ?it/s]
Predicting DataLoader 0:  50%|          | 1/2 [00:02<00:02,  0.34it/s]
Predicting DataLoader 0: 100%|          | 2/2 [00:03<00:00,  0.58it/s]
Predicting DataLoader 0: 100%|          | 2/2 [00:04<00:00,  0.49it/s]
```





12

Not just on SLURM @ INFN and ICSC

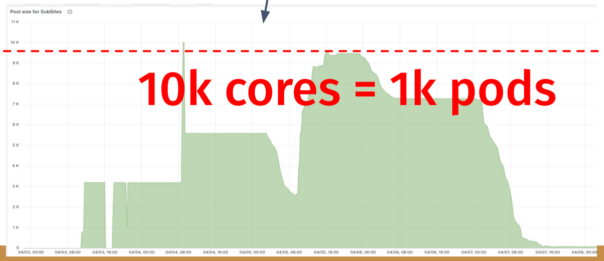
Integrated HTC resources from Tier2 into the [INFN "Analysis Facility".](#)
Now in production → users can spawn **Dask clusters seamlessly on every Italian Tier2.**

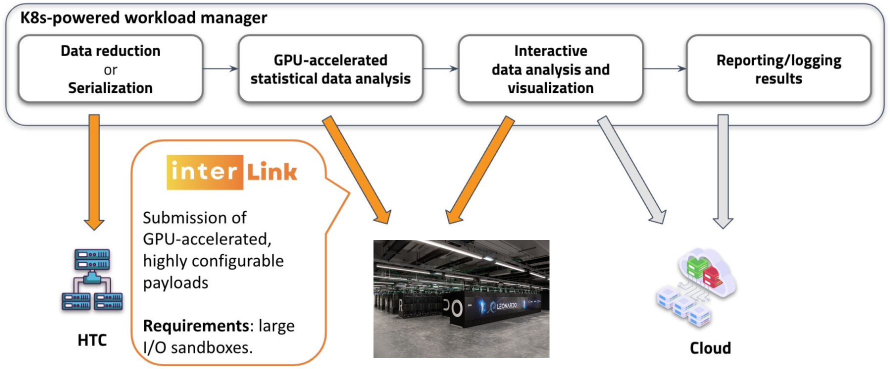
AI_INFN initiative have built **a prototype on top of interLink** (**[Matteo's talk](#)**)
Giving ML/AI experts access to **GPUs** via **a single interface:**

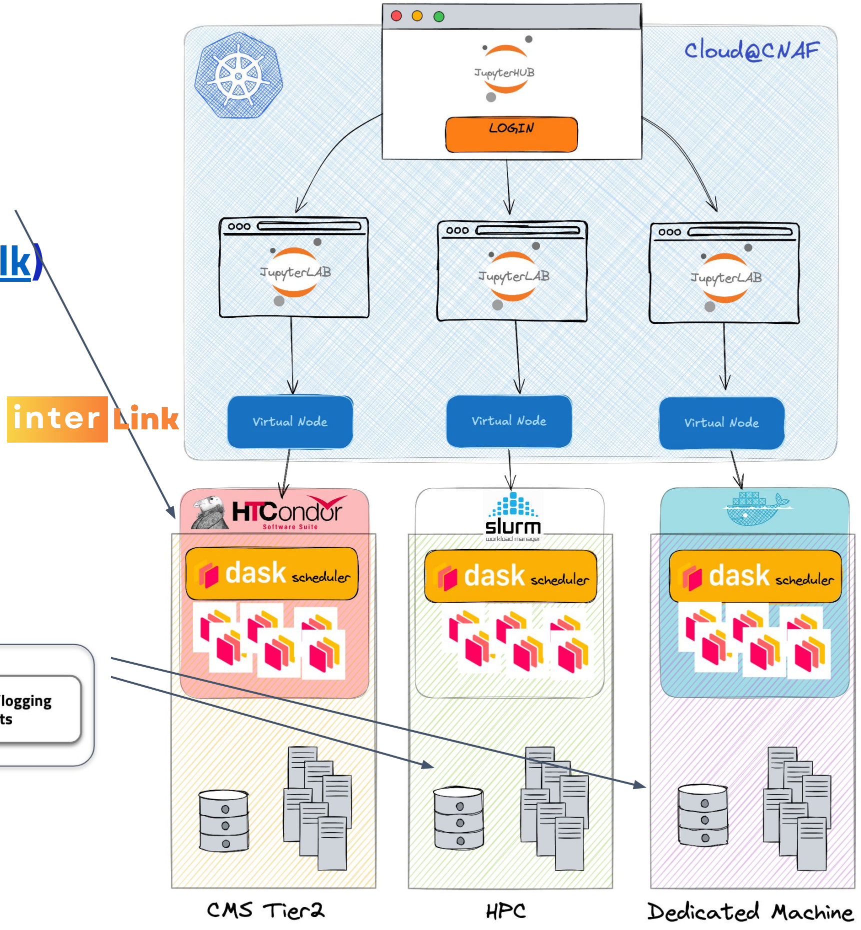
- **on LEONARDO@CINECA**
- on **beefy GPU equipped VM**

ICSC PoC for cloud and Leonardo integration
[See Claudio's talk](#)

Deploying CMS pilot job at scale?
[See Daniele's talk](#)





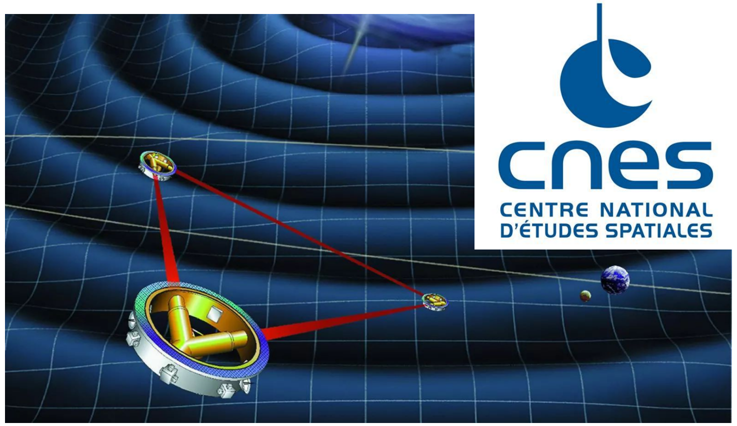


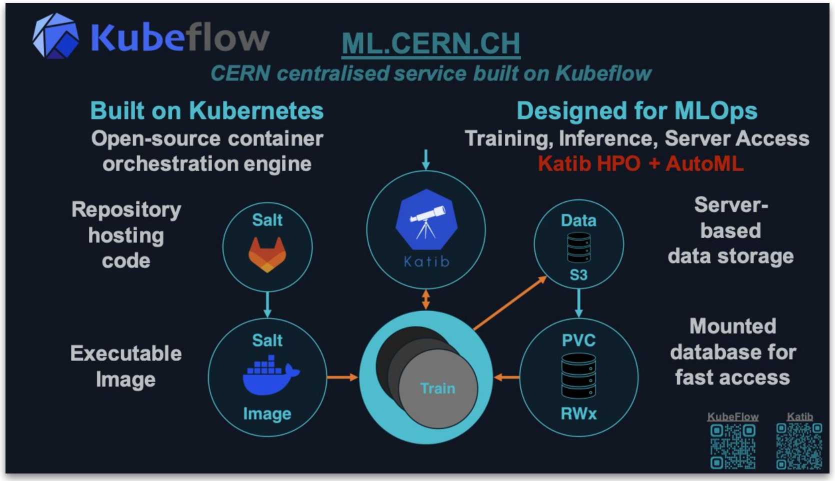
Toward a "community" software

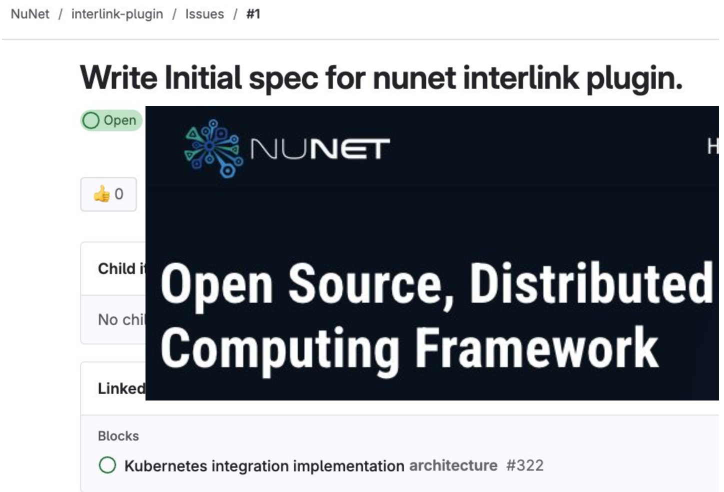
ArgoWorflows from LISA **interferometer** are being evaluated to offload argo tasks to HPC centers running SLURM (INFN-CNES interactions)

We are in implementation phase for a offloading prototype with **CERN KubeFlow** based platform.

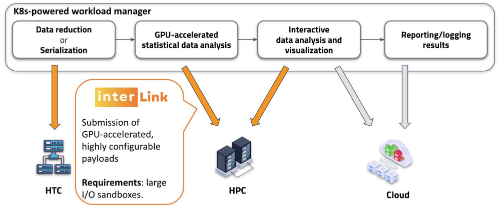
"Non-scientific" case studies are also joining the evaluation with their own plugins **(NuNet)**





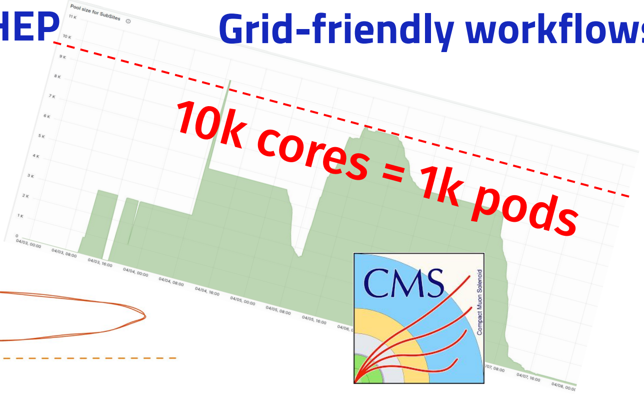


HOW IT'S GOING? Not too bad!

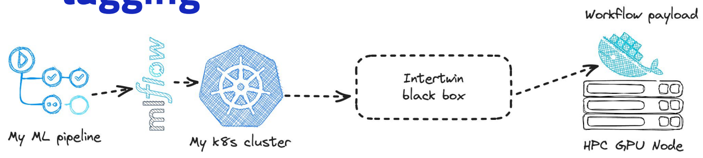


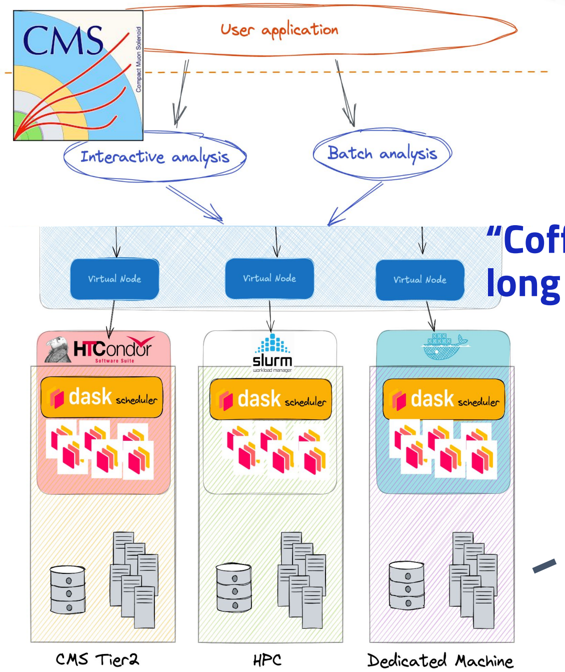
GPU accelerated HEP simulations and triggers

Grid-friendly workflows

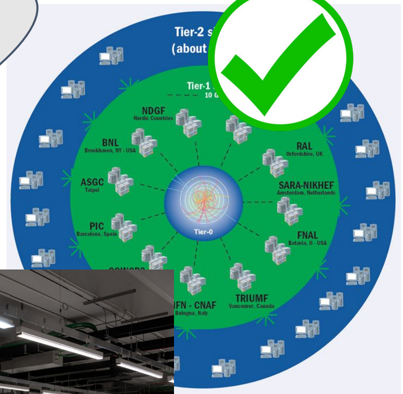


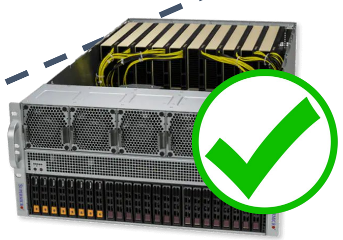
ML training and inference of particle tagging





"Coffee-break-duration" long full analyses at LHC





GPU cloud machine

Summary

We demonstrated how interLink can extend our resource provisioning model...

for ANY container-based use case!

HPC-Slurm is only the beginning

The inclusion of "any" kind of remote resource can work (CaaS, fat nodes, HTC etc..)

Providers can offer an effortless K8s style access to resources

without internals deep knowledge required, develop your own provider plugin!

Next steps:

Streamline communities onboarding

documentation and development SDKs

Consolidate through use cases feedback loop

community driven development is key

Widening community adoption

applying for contributing to Cloud Native Computing Foundation Sandbox interLink

https://intertwin-eu.github.io/interLink/

Future Artificial Intelligence Research

This work is also partially supported by the FAIR "Future Artificial Intelligence Research" NRRP Project

BACKUP