



interTwin

Il progetto interTwin e le sue sinergie tecnologiche per l'INFN

Daniele Spiga a nome di... (vedi dopo)



interTwin is funded by Horizon Europe under grant agreement n° 101058386

- **Cosa è interTwin e l'INFN**
 - Numeri e responsabilità
- **Overview dell'architettura**
 - E focus su integrazione risorse eterogenee
- **Sintesi e sinergie tecnologiche per l'INFN**

Marica Antonacci

(INFN-BA): Task Lead 5.4

Sara Vallero

(INFN-TO): Task Lead 4.1

Alberto Gennai

(INFN-PI): Task Lead 7.3

Diego Ciangottini - in sostituzione

(INFN-PG): Task Lead 5.1

Daniele Spiga

(INFN-PG) WP5 Lead

Davide Salomoni, Tommaso Boccali, Massimiliano Razzano, Federica Legger, Luciano Gaido, Alessandro Costantini, Stefano Bagnasco, Giancarlo Cella, Giacinto Donvito, Giorgia Minnello, Ahmad Alkhansa, Giacomo Surace, Luca Giommi.... *<mi scuso se ho dimenticato qualche nome>*

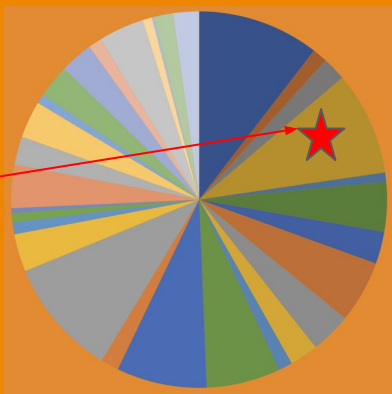
INFN-BA, INFN-CNAF, INFN-PG, INFN-PI, INFN-TO



Durata 36 Mesi
da
Settembre 2022
a
Agosto 2025

30 Partners
Con
1500 PM totali
Coordinato da
EGI

Budget 12M Euro
INFN 1M Euro



2 + 2 + 3 DT
Use Cases

dalle comunità

HEP
Astro / GW
Climate
Environment



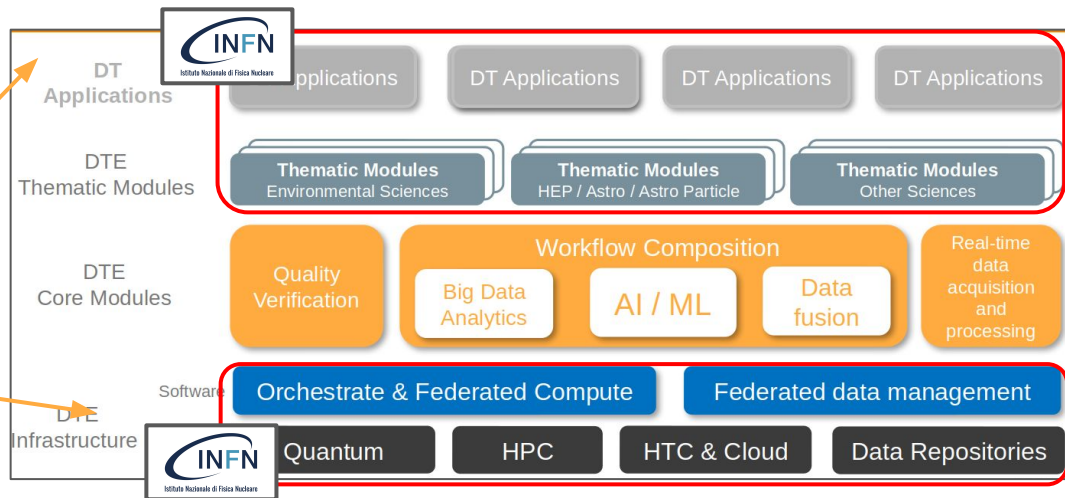
Cosa fa il progetto e i contributi INFN



Creazione di un prototipo di Digital Twin Engine (DTE):

- DTE Blueprint Architecture
 - Implementazione di un prototipo
- INFN contribuisce su due assets**
- Use Case VIRGO/ET
 - Infrastruttura del DTE

Sviluppi/Integrazioni infrastrutturali li possiamo testare/guidare con lo use case scientifico che abbiamo in casa



Gemello digitale
o
Simulazione ??



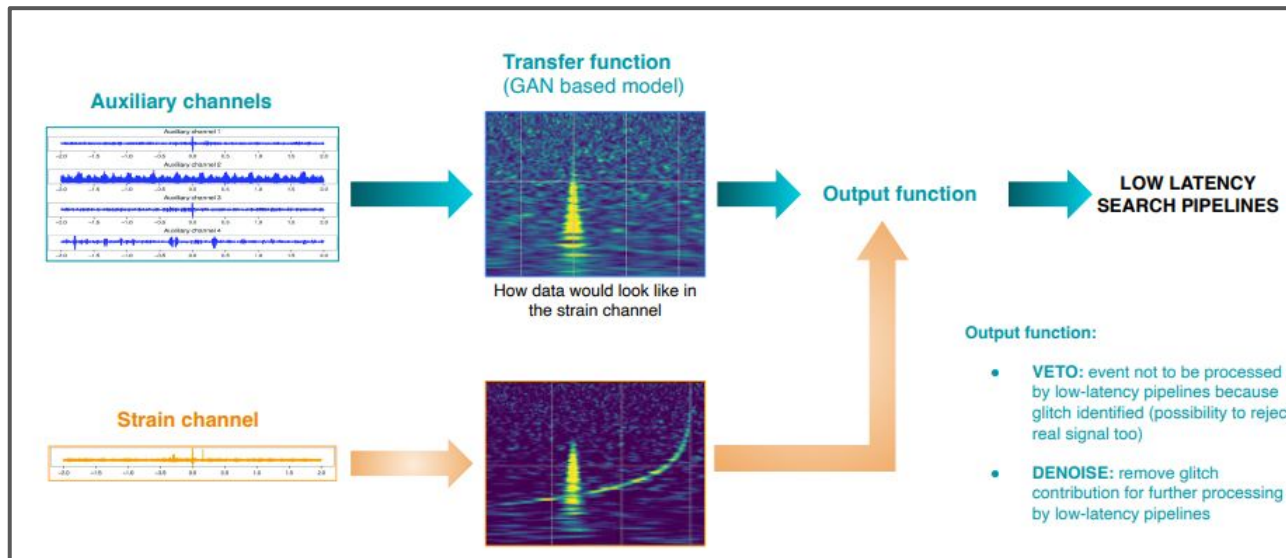
A digital twin is a **virtual replica or simulation of a physical object**, system, or process, created and maintained using real-time data, advanced analytics, and machine learning algorithms. This digital representation enables continuous monitoring, analysis, and optimization of its real-world counterpart, facilitating better decision-making, improved efficiency, and reduced operational risks. Digital twins are widely used in various industries such as **manufacturing, transportation, energy, and healthcare for predictive maintenance, product design, and performance optimization.**



Ha lo scopo di simulare realisticamente il rumore nel rivelatore, al fine di:

- studiare come reagisce ai disturbi esterni
- **in prospettiva di Einstein Telescope** essere in grado di rilevare i "glitch" di rumore in tempo quasi reale (attualmente non possibile).

Il risultato sarà quello di riuscire ad inviare trigger più affidabili agli osservatori per la multi-messenger astronomy.





L'infrastruttura del DTE

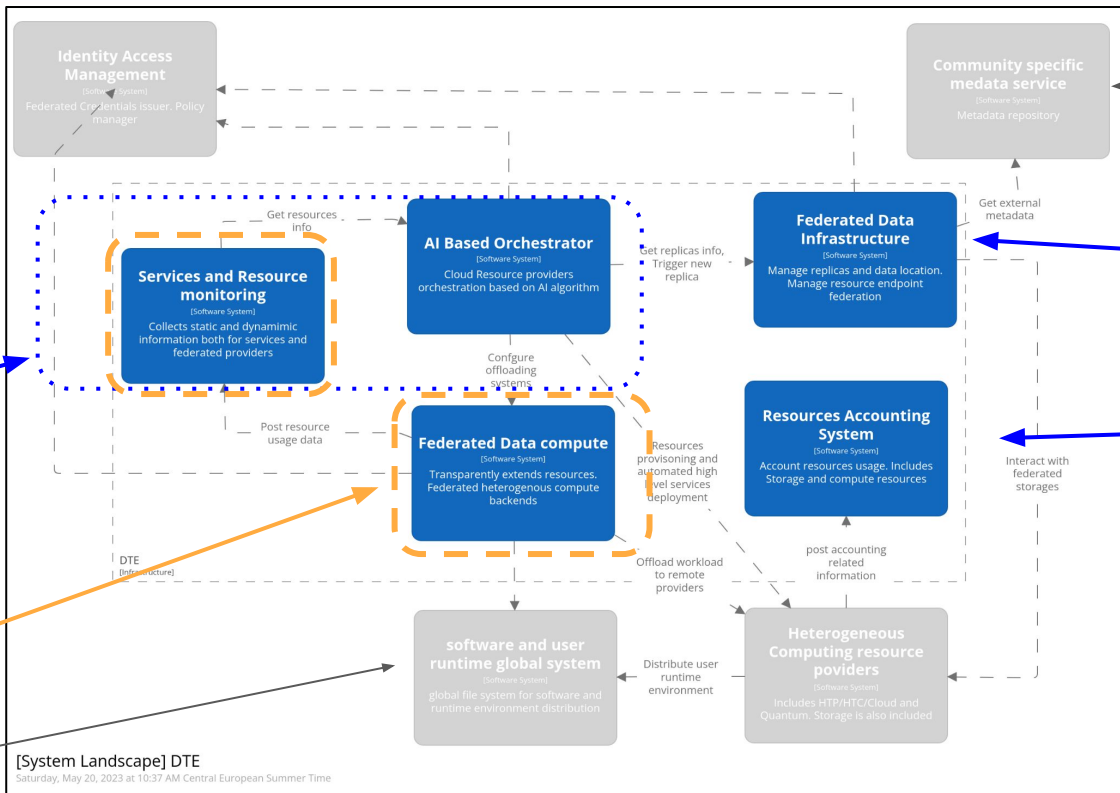
INDIGO-IAM
oppure EGI Check-IN

Ovvero quell'insieme di servizi che da accesso alle risorse
Non partiamo da zero!! Ma evolviamo quanto già fatto

INDIGO PaaS
Orchestrator

Sistema per l'offloading
(vedi dopo)

CVMFS



[System Landscape] DTE
Saturday, May 20, 2023 at 10:37 AM Central European Summer Time



I servizi

Già nel proposal abbiamo spiegato come avremmo potenziato (quando possibile) i servizi esistenti in particolare:

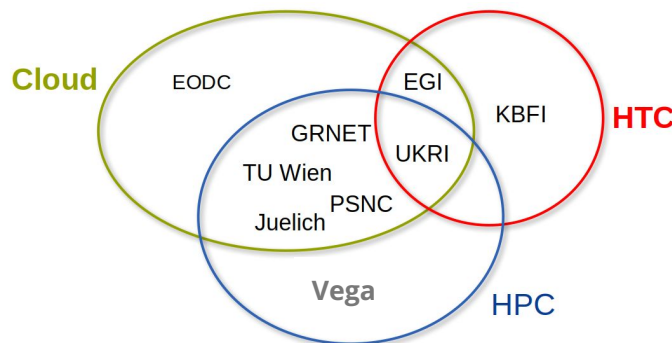
- Orchestrazione Cloud e deployment di servizi on demand su modello DataCloud ([vedi Talk Marica](#))
 - Evoluzione della logica di selezione dei providers
 - Integrazione con modello di offloading (vedi dopo)
- Datalake implementato a la WLCG ovvero a la Escape (Vedi [Talk Diego](#) & [Demo Massimo](#))
 - Estensione verso HPC, Caches etc
- **Integrazione di centri HPC e risorse eterogenee: offloading**
 - Evoluzione tecnologica dell'estensione trasparente Tier in WLCG (e.g. T1-CNAF su CINECA etc)

Providers di Risorse

I providers che partecipano al consorzio non ricevono effort per fornire risorse ma per fare attività di integrazione

- Elemento molto importante
- Sono partners di WP5 (coordinato da INFN)

I providers sono stati selezionati per rappresentare uno scenario significativo

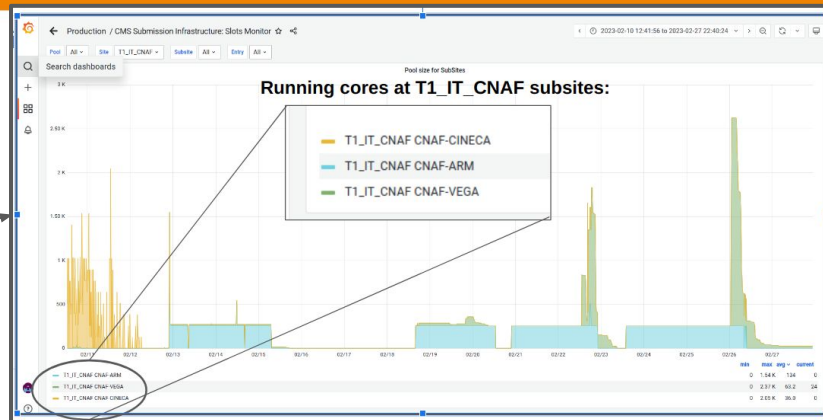




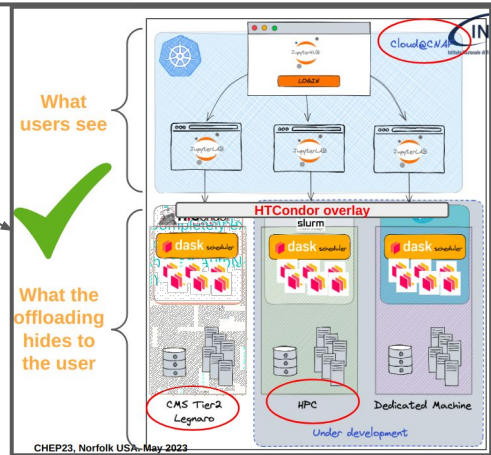
Focus su modello di offloading

Elemento nuovo nell'architettura proposta. Il modello ha origine da alcune esperienze pregresse che **hanno funzionato e stanno funzionando**. In particolare:

- **Estensione trasparente del Tier1 su risorse opportunistiche**
 - CNAF e CINECA (A2, M100 Galileo), Risorse opportunistiche & CNAF Reloaded
- **Integrazione trasparente di risorse INFN-Cloud e Grid (Tier2) per analisi dati ad alto throughput**
 - Analysis Facility @ CMS
 - Attività di Spoke2



Vogliamo generalizzare il modello, strutturarlo e renderlo facilmente riutilizzabile. Ovvero, ad esempio, un possibile sistema di integrazione di risorse eterogenee (i.e. Cloud/Grid) “@ INFN & beyond”



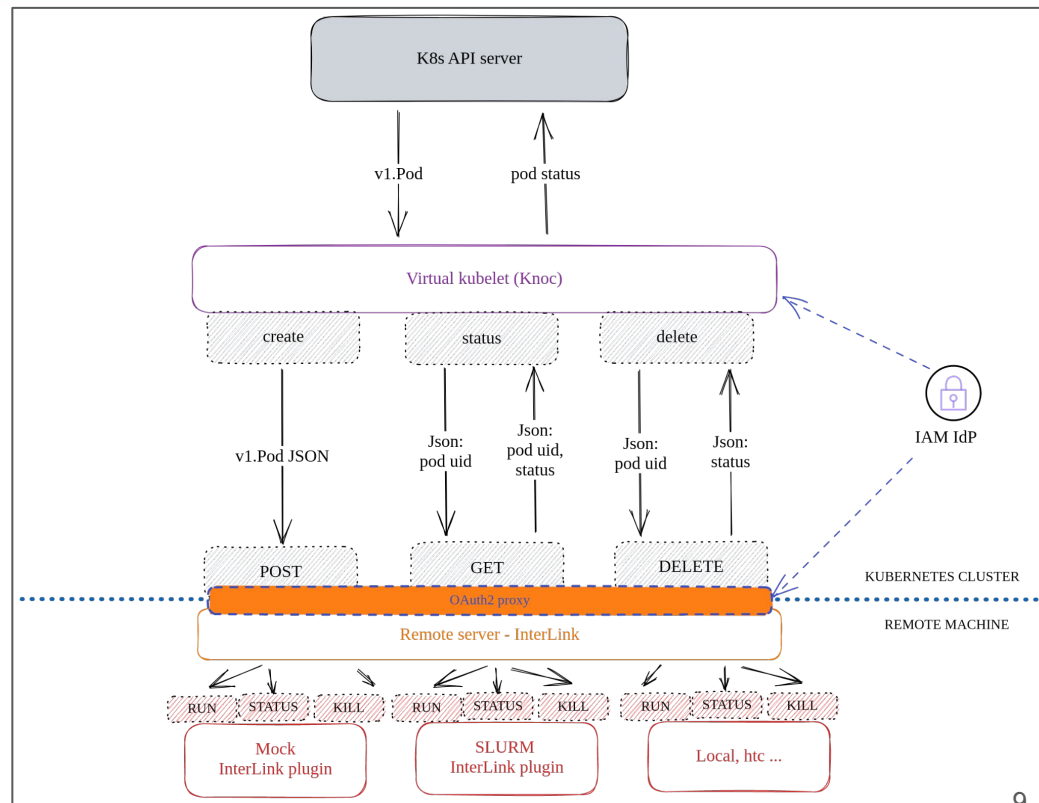
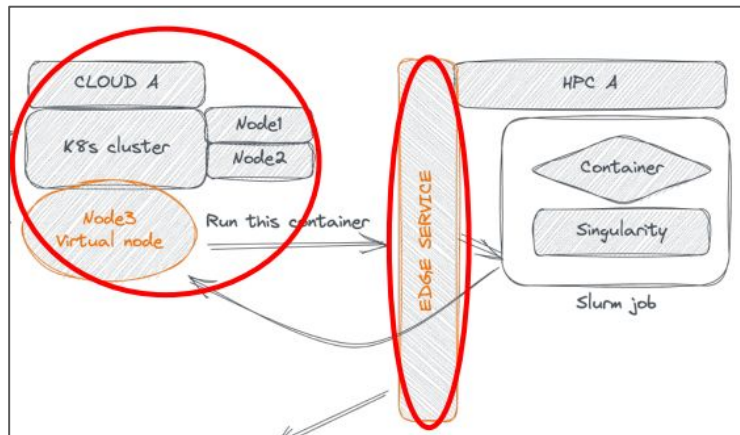


Come proponiamo di farlo

Usando **Virtual Kubelet** (estendere k8s senza imporre dipendenze sw specifiche di k8s)

Sviluppando un **layer API** da schierare in prossimità (edge) di una "qualunque risorsa" di calcolo.

Assunzioni principali: CVMFS accesso nodo - storage per dati (Datalake)





L'offloading in pratica (Jupyter è solo un esempio)



1

Server Options

Select your desired image: ghcr.io/dodas-ts/htc-dask

Select your desired number of cores: 1

Select your desired memory size: 2GB

Enable Offloading to: **Vega GPU**

Start

2

hub.131.154.96.26.myip.cloud.infn.it/hub/spawn-pending/spiga

Your server is starting up.
You will be redirected automatically when it's ready for you.

2023-05-20T09:17:26.656919Z [Normal] Successfully assigned jhub/jupyter-spiga to veiga-v...

```

drwxr-xr-x 50 dantele dantele 135168 mag 20 10:31
dantele@icarbon:~/Downloads$ rm -rf structurizr-1-
structurizr-1-D1a2.png          structurizr-1-system_view_full.png
structurizr-1-D104.png         structurizr-1-system_view.png
dantele@icarbon:~/Downloads$ rm -rf structurizr-1-*
dantele@icarbon:~/Downloads$ ssh spigad@vgl01n0001.veiga.lzum.si

WELCOME TO VEGA

You are logging on to the equipment of the Vega Cluster
Login and use of equipment by unauthorized personnel is strictly prohibited!
All connections are monitored and recorded.
Disconnect IMMEDIATELY if you are not an authorized user!

Last failed login: Fri May 19 12:38:19 CEST 2023 from 139.186.16.241 on ssh:notty
There were 6 failed login attempts since the last successful login.
Last login: Thu May 18 16:43:23 2023 from 141.236.255.10
[spigad@vgl01n0001 ~]$ squeue -me
          20910 PARTITION  NAME          USER ST      TIME  NODES NODELIST(REASON)
[spigad@vgl01n0001 ~]$ squeue -me
          62758880      gpu testvk... spigad  R      0:05   1 gn07
  
```

3

jhub.131.154.96.26.myip.cloud.infn.it/user/spiga/lab/tree/Untitled.ipynb

```

[1]: [nvdi...
Sat May 20 11:21:43 2023
-----
NVIDIA-SMI 530.30.02                Driver Version: 530.30.02  CUDA Version: 12.1
-----+-----
GPU Name          Persistence-M  Bus-Id        Disp.A    Volatile Uncorr. ECC
Fan   Temp   Perf          Pwr:Usage/Cap      Memory-Usage  GPU-Util  Compute M.
-----+-----+-----
0  NVIDIA A100-SXM4-40GB      On           00000000:03:00:0  Off          0
N/A    44C    P0              56W / 400W      6MiB / 4896MiB    0%         Default
                                     Disabled
-----+-----
Processes:
GPU   GI   CI   PID  Type  Process name                      GPU Memory
ID   ID                                     Usage
-----+-----
No running processes found
  
```

- 1 JupyterHub @ CNAF (su k8s)
- 2 Jupyter Lab @ Vega (via slurm)
- 3 Accesso user-level a GPU (V100) disponibile su WN a Vega

IP Details for: 131.154.96.26

Decimal: 2207932442

Hostname: cloud-vm26.cloud.infn.it

ASN: 131

ISP: INFN - CNAF - Bologna

Services: None detected

Assignment: Libby Static IP

Country: Italy

State/Region: Emilia-Romagna

City: Bologna

IP Details for: 1535.68.41

Decimal: 2567259177

Hostname: vgl01n0001.veiga.lzum.si

ASN: 2107

ISP: Institut Informacjnyh Znanosti Maribor

Services: None detected

Assignment: Libby Static IP

Country: Slovenia

State/Region: Maribor



Abbiamo definito l'architettura dell'infrastruttura, i workflow e i requirements degli use cases,

- F2F technical meeting a Marzo 2023

Abbiamo avviato l'attività dei pilots

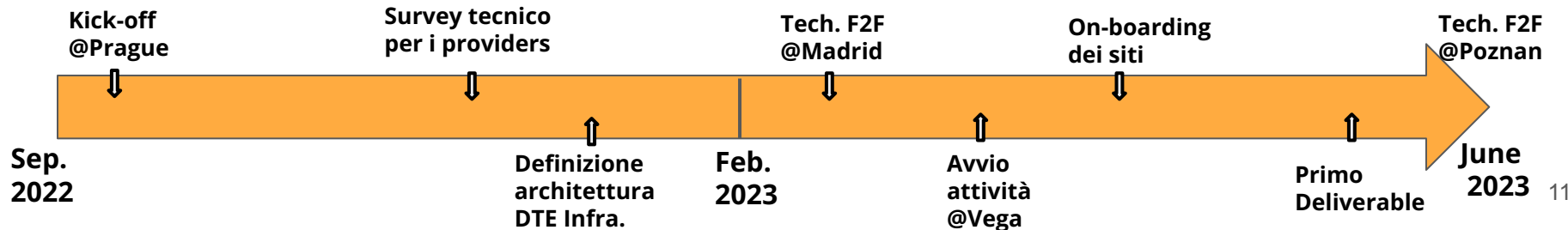
- Questa è una priorità per noi e per il progetto in generale.
- Abbiamo attivato 4 siti per il Datalake (DESY, GRNET+CERN e **INFN-CNAF**)
- Abbiamo 4 siti per il Compute (Vega, Juliech, UKRI + **INFN-Cloud**)

In parallelo ci coordiniamo internamente @INFN



April 2023	
18 Apr	Third interTwin - INFN meeting
January 2023	
16 Jan	Second interTwin - INFN meeting
December 2022	
16 Dec	First interTwin - INFN meeting

Stiamo preparando il secondo technical meeting a Giugno 2023





In sintesi: Le sinergie per l'INFN

Principali sviluppi tecnologici interTwin (in WP5)

Attività CORE e progetti INFN

	Sistema di Offloading	Orchestrazione Cloud	DataLake Rucio based Data Management	CVMFS
DataCloud & ICSC Spoke0		Servizi alto livello (deployment on demand) WP5	Sistema pilota per il Data Management Nazionale [Talk qui], [Demo qui]	Sistema pilota per la gestione software (e dei container) [Demo qui]
TeRABIT	Integrazione BubbleHPC			
ICSC Spoke2 /3	Analisi Alto throughput (via Dask)	Deployment di servizi comunità/affiliati	Datalake per attività scientifica e grant industriali (i.e. Leonardo)	Soluzione per distribuzione immagini (e software)
CMS	Analysis Facility (analisi ad alto throughput) [Talk qui] Integrazione HPC			

N.B: questo schema è relativo esclusivamente alla parte tecnologica di interTwin e non quella scientifica



Il progetto è stato avviato e ci sono molte sinergie con le attività core dell'INFN. Io non vedo criticità.

- Il primo giro di Deliverables è sostanzialmente concluso

L'aspetto più urgente/critico è quello di avviare l'**integrazione con gli use cases**

- Implementare i pilots rappresenta una importante motivazione per provare e dare feedback
- **Oltre Virgo/ET, siamo in stretto con CERN** che a loro volta sono interessati ai nostri sviluppi per l'offloading

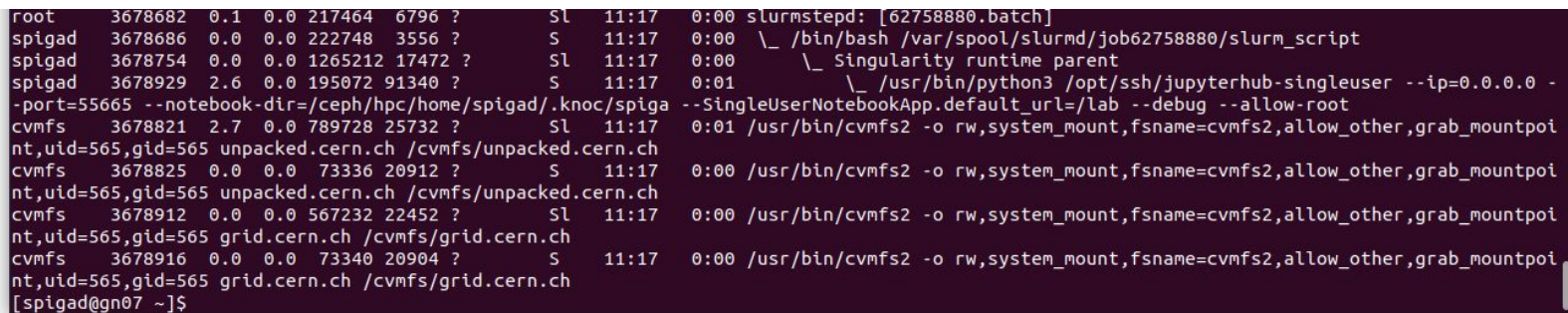
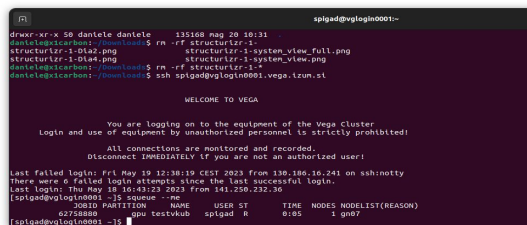
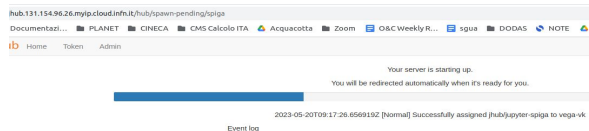
La partecipazione attiva di (alcuni) provider rende l'attività molto interessante (divertente) e soprattutto **permette di fare lavoro!**

- **Molto importante il supporto infrastrutturale dell'INFN (CNAF & CLOUD)!! E' molto apprezzato e utile**

Rendicontazione STAFF INFN, burocrazia di progetto (di qualunque tipo), tutto MOLTO faticoso... il carico PNRR sulle persone non aiuta.

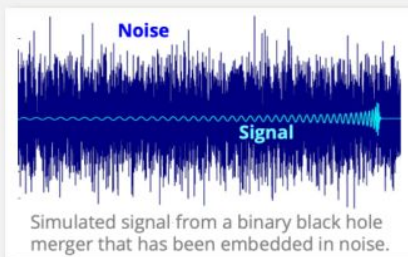
Backup





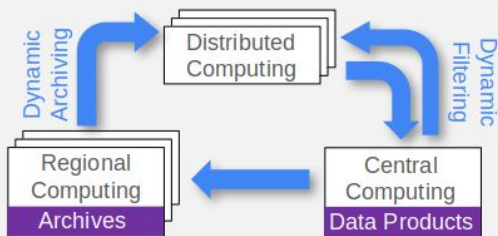


DT of the VIRGO Interferometer



It is meant to **realistically simulate** the noise in the detector, in order to study how it reacts to external disturbances and, in the perspective of the **Einstein Telescope**, to be able to detect noise “glitches” in **quasi-real time**, which is currently not possible. This will allow sending out **more reliable triggers** to observatories for multi-messenger astronomy.

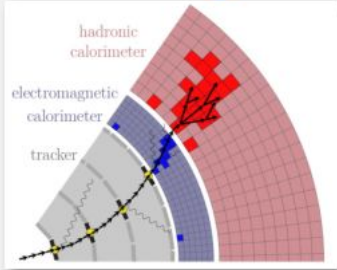
DT for noise simulation of next-generation radio telescopes



Meant to provide DTs to simulate the noise background of radio telescopes (**MeerKat**) will support the identification of rare astrophysical signals in (near-)real time. The result will contribute to a realisation of "**dynamic filtering**" (i.e. steering the control system of telescopes in real-time).

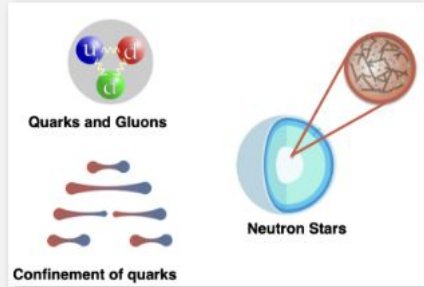


DT of Large Hadron Collider (LHC) detector components



Seeking for strategies to face the increase in the need for simulated data expected during the future High Luminosity LHC runs. The primary goal is to provide a fast simulation solution to complement the Monte Carlo approach. ***Faster and deeper cycles of optimisation of the experiment parameters*** in turn will enable breakthroughs in experimental design.

DT of the Standard Model in particle physics

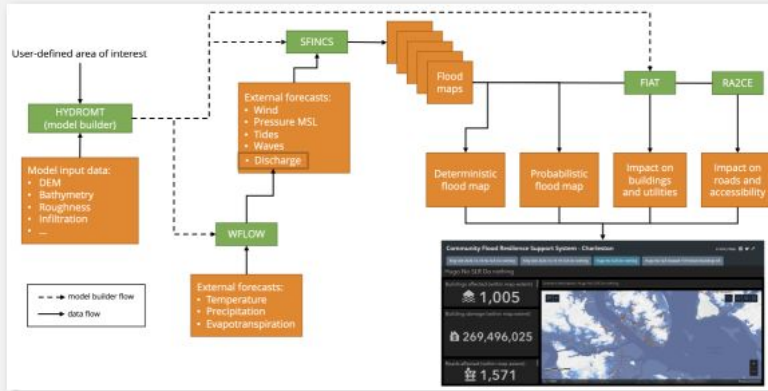


Competitive results in Lattice QCD require the ***efficient handling of Petabytes of data***, therefore the implementation of advanced data management tools is mandatory. On the side of algorithmic advancement, ML algorithms have recently started to be applied in Lattice QCD. The goal is to ***systematize the inclusion of ML for large scale parallel simulations***.

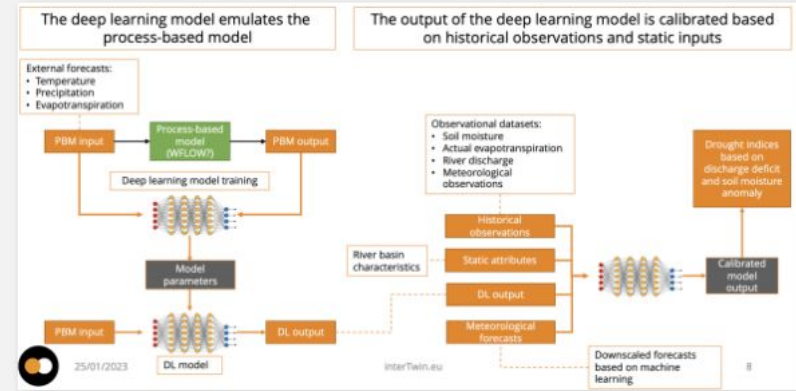


DT for Extreme Events on the Earth

- Climate Change Future Projections of Extreme Events as **storms and fire**;
- Early Warning for Extreme Events of **floods and droughts**;
- Climate Change Impacts of Extreme Events of **storms, fire, floods and droughts**.



Digital Twin for Flood Early Warning in coastal and inland regions



Digital Twin for Drought Early Warning in alpine regions



WP5 Allocated effort

WP5 – DT Engine Infrastructure (Lead beneficiary: INFN) [M1-M36]								
Part. no	1	2	4	5	10	12	15	16
Short name	EGI.eu	AGH	CERN	CESNET	DESY	EODC	GRNET	INFN
PMs /part.	4	19	24	12	36	8	12	76
Part. no	17	18	19	24	25	26	30	31
Short name	FZJ	KIT	KBFI	PSNC	TU Wien	UKRI	JSI	IZUM
PMs /part.	12	15	12	12	8	33	12	3



Total PMs 298
Total N° Partners: 16



Four Tasks

- **TT5.1. Federated compute infrastructure (HTC, HPC, Cloud, Quantum)**
 - **Lead: Diego Ciangottini [INFN]**
 - **The main goal is to provide software solutions to enable resources provisioning on a wide range of compute providers**
- **T5.2 Federated data infrastructure**
 - **Lead: Paul Millar [DESY]**
 - **The main goal is to support data requirements of digital Twins**
- **T5.3. Federation Services and policies**
 - **Lead: Ian Neilson [UKRI]**
 - **The main goals of T5.3 are security and accountability for users and providers by creating**
- **T5.4 AI-based orchestrator**
 - **Lead: Marica Antonacci [INFN]**
 - **The main goal is to orchestrate data and compute resources "intelligently", taking into account data location and performance metrics.**



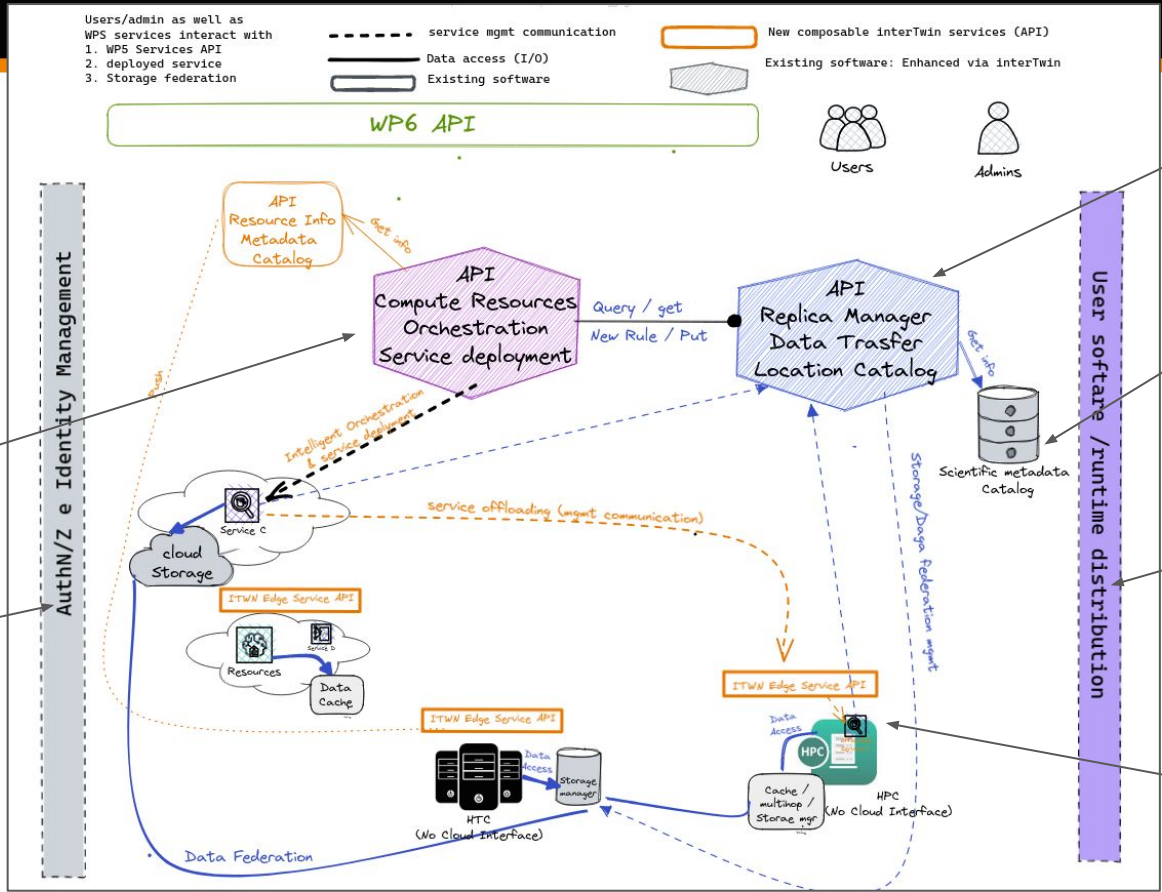
L'infrastruttura del DTE

Ovvero quell'insieme di servizi che da accesso alle risorse

Non partiamo da zero!! Ma evolviamo quanto già fatto

INDIGO PaaS Orchestrator

INDIGO-IAM Oppure EGI Check-IN



Rucio + FTS

Cataloghi di comunità

CVMFS

Sistema per l'offloading (vedi dopo)

Thank you!

