Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

INFN

CTlab4ET
Computing Technology Laboratory for Einstein Telescope

Luca Tabasso,
Stefano Bagnasco, Lia Lavezzi, Federica Legger, Stefano Lusso, Nicola Mosco, Sara Vallero

EINSTEIN TELESCOPE

INFN TORINO

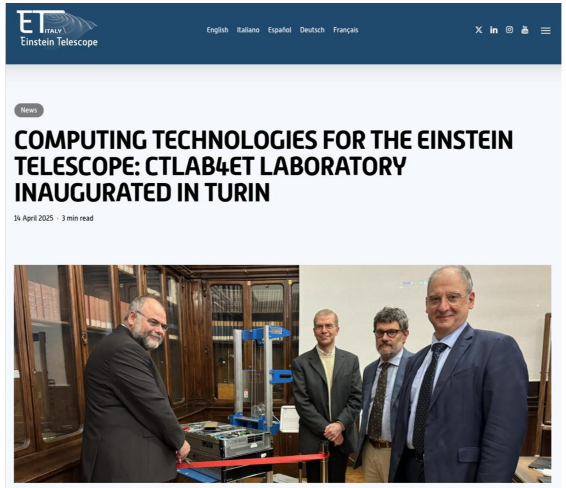
ET ITALY Einstein Telescope

funded by the project ETIC: Einstein Telescope Infrastructure Consortium (ETIC - IR0000004)
PNRR MISSIONE 4, COMPONENTE 2, INVESTIMENTO 3.1

INFN-Torino CTlab4ET

- INFN Torino ha inaugurato un nuovo laboratorio **dedicato a ET**:
  "**CTLab4ET (Computing Technology Laboratory for ET)**"
  un cluster eterogeneo orientato alle attività di computing R&D e di supporto al Technology Tracking per la definizione del Computing Model di ET, il cui hardware è stato soprannominato **TechZoo**
- **Scopo**
  - Porting del codice su architetture differenti da x86 (GPU, Arm, ecc…)
  - Test su configurazioni eterogenee
  - Benchmark e valutazione di architetture in funzione della sostenibilità (carbon footprint, ecc…)
  - Computazione parallela e HPC in un ambiente completamente controllato e customizzabile
  - Supporto allo sviluppo di middleware e framework per ET
  - Supporto alle esigenze di calcolo per la caratterizzazione del sito

Inaugurazione 14 Aprile 2025



www.einstein-telescope.it

www.infn.it

TechZoo animals

Potenza di calcolo, finanziamento ETIC

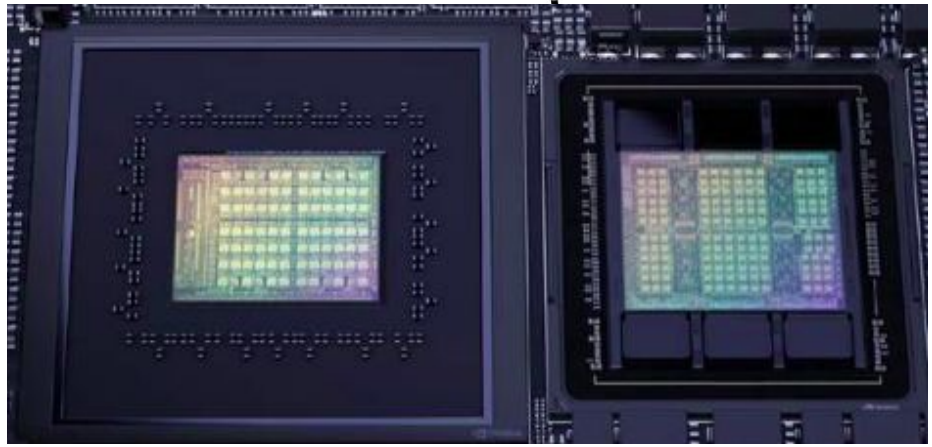
- 4× Dell PowerEdge R7525
  - 2× AMD EPYC 7313 3.0 GHz (16 cores)
  - 1TB RDIMM, 3200 MT/s
  - 2× PCIe Gen4 slots + 2× PCIe Gen4 GPU slot DW
  - 2× SSD 960 GB with Endurance DWPD ≥ 0.9
- 3× Lenovo
  - 2× AMD EPYC 9124 3.0 GHz (16 cores)
  - 1TB RAM (expandable)
  - PCIe Gen5 slots
  - InfiniBand connectivity and NVIDIA ConnectX-7 NDR OSFP400

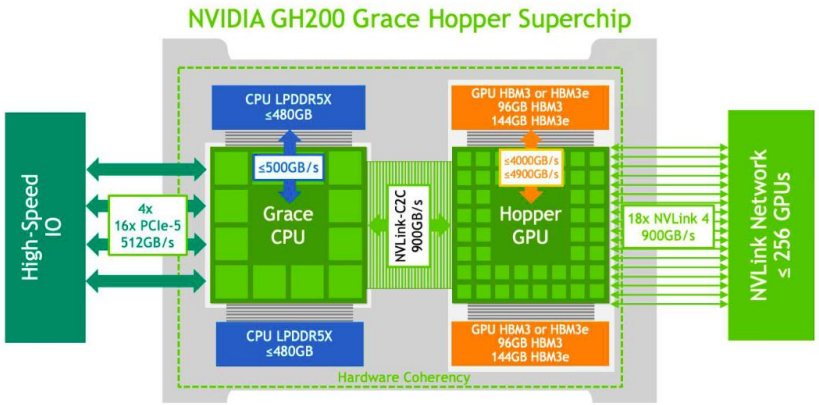
- GPUs DW
  - 2× AMD MI100, 32GB
  - 1× NVIDIA A16, 4×16GB
  - 2× NVIDIA L40S 48GB
  - 2× NVIDIA A5000, 24GB

TechZoo *odd* animal

Acquistato come sistema di valutazione, finanziamento ETIC.

- 1× NVIDIA Grace Hopper Superchip
- CPU+GPU a modello di memoria coerente
- 900 GB/s interfaccia coerente NVLink-C2C
- Adottato nell'Alps Supercomputer @ CSCS, [Swiss National Supercomputing Center]





- **CPU progettata per applicazioni AI e HPC**
- **7x bandwidth rispetto x16 PCIe Gen5**
- **I thread CPU e GPU threads possono ora accedere contemporaneamente e in modo trasparente sia alla memoria CPU che quella GPU**

https://www.nvidia.com/en-us/data-center/grace-hopper-superchip

TeRABIT Interplay

Il progetto **TeRABIT** persegue due obiettivi principali:

- Potenziare l'infrastruttura di rete nazionale italiana per la ricerca e l'istruzione, fornendo capacità di connessione fino a un Terabit al secondo (1 Tbps).
- Realizzare sistemi di Calcolo ad Alte Prestazioni (HPC) distribuiti geograficamente, denominati "HPC Bubbles", progettati per offrire risorse HPC delocalizzate e vicine agli utenti finali.

- Il nodo del cluster TeRABIT di Torino, è configurato con le seguenti specifiche hardware:
  - 6 CPU nodes lenovo: 2x(AMD EPYC 9654 96-Cores), 1,5TB RAM, InfiniBand NDR 400Gbps per node
  - 6 GPU nodes lenovo: 2x(AMD EPYC 9654 96-Cores), 1,5TB RAM, InfiniBand NDR 400Gbps per node
    - NVIDIA H100 SXM5 80GB, HBM3 memory per node
  - 1 InfiniBand switch 400 Gb/s

Possibilità di integrazione in INFN Data Cloud

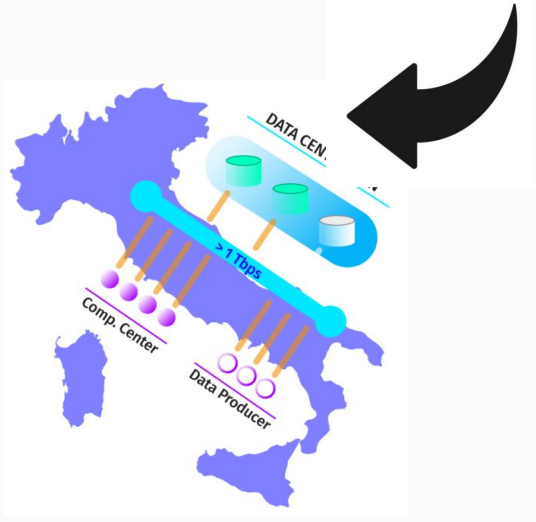
ETIC "TechZoo" Cluster for R&D

Modular

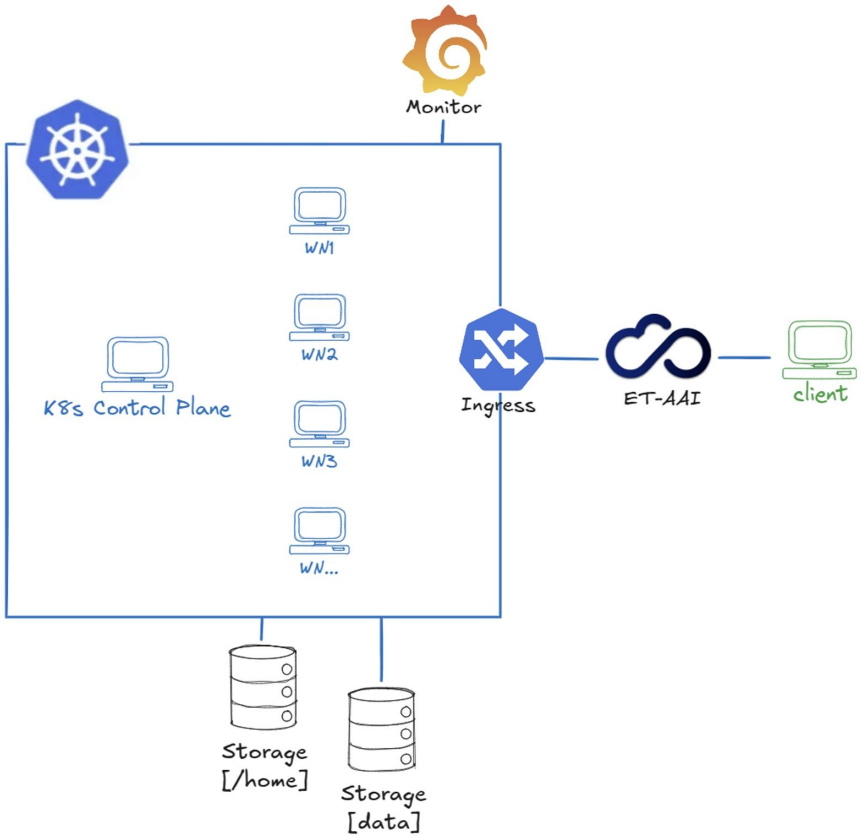
Infiniband

INFN Data Cloud

TeRABIT Cluster
Terabit network for Research and Academic Big data in ITaly



CTlab4ET Design



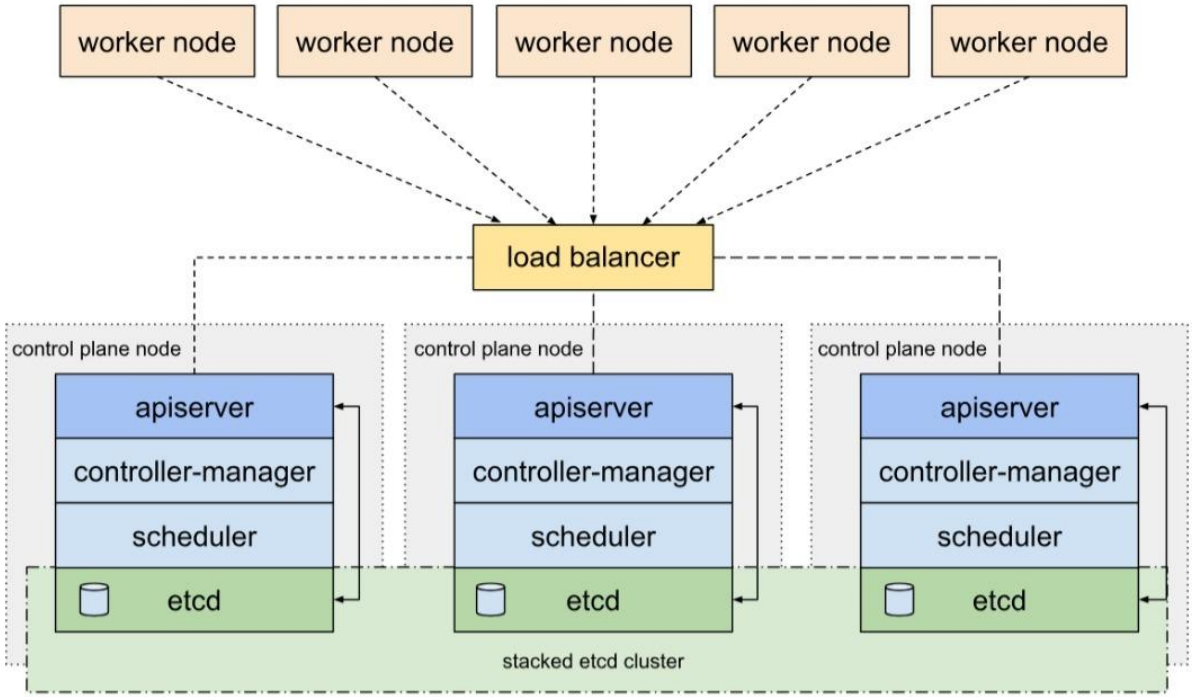
Cluster Kubernetes (Bare-Metal)

- Installazione in High Availability
- Storage separato per /home utenti e data
- Autenticazione via ET-IAM
- Monitoring
- Molteplici servizi implementati sotto forma di pod

Stato dei lavori: Control Plane

Control Plane in High Availability:

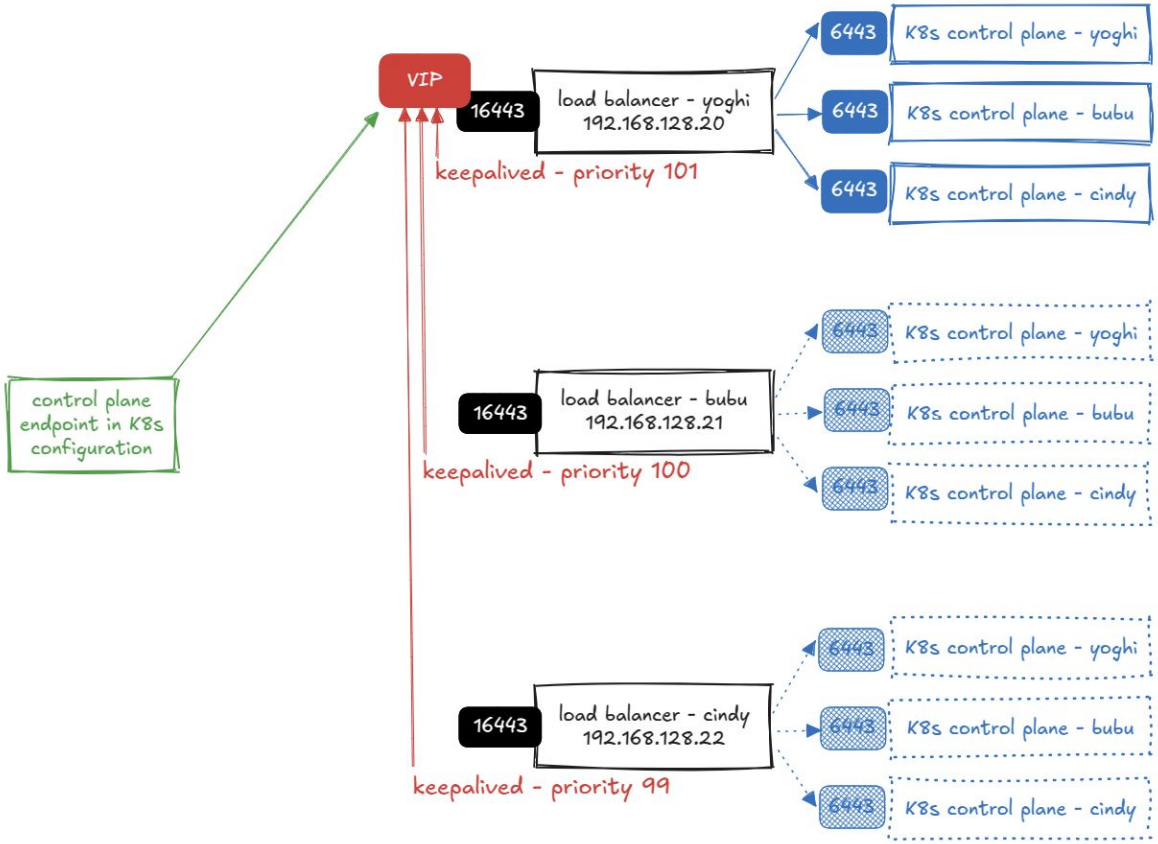
- Topologia stacked etcd;
  - Copia etcd db in ogni nodo
- HAProxy come load-balancer
  - Riceve richieste per apiserver e le ridistribuisce
  - algoritmo `leastconn`: viene scelto il backend col minor numero di connessioni attive
  - Health-checks: se un API server non risponde correttamente, HAProxy smette di inviargli traffico.



Stato dei lavori: Load Balancer

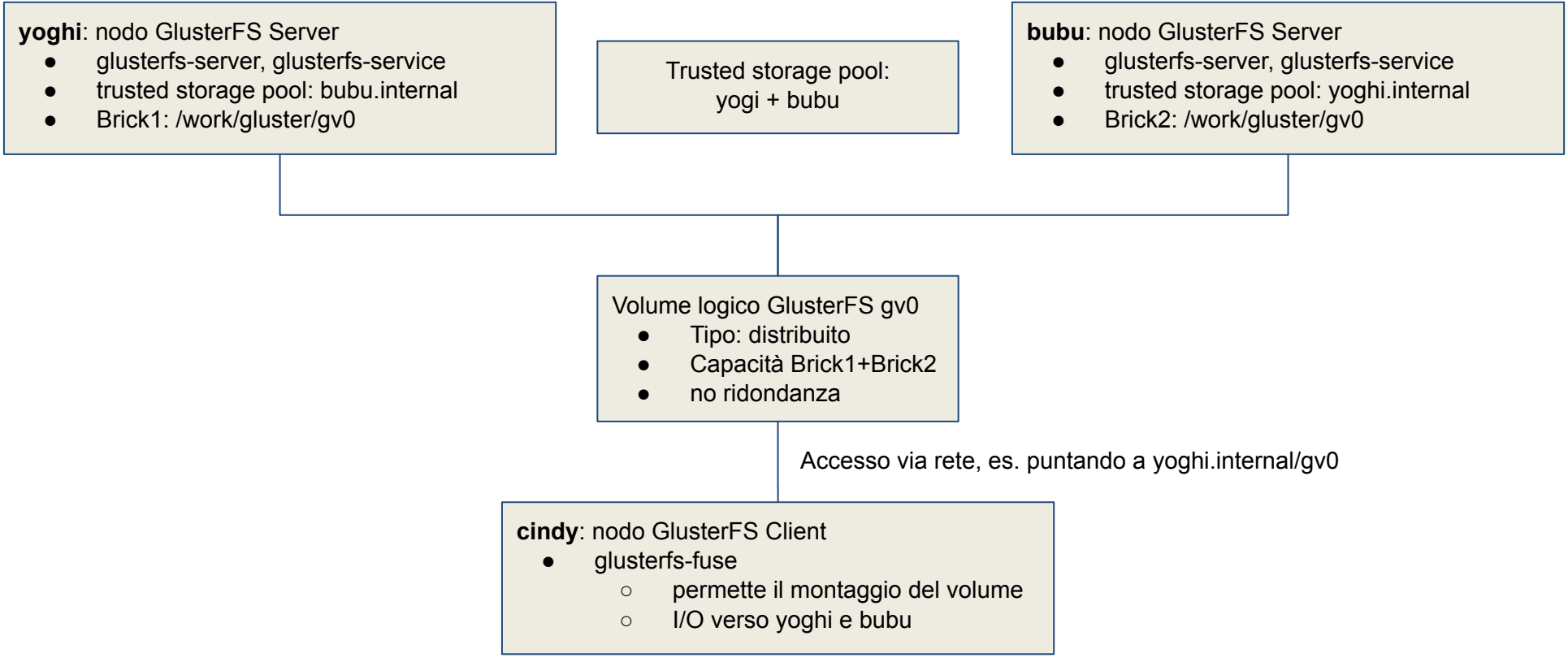
Load Balancer in High Availability:

- Keepalived+VIP (Virtual IP address):
  - usando VRRP (Virtual Router Redundancy Protocol) le tre macchine condividono un VIP, usato attivamente da quella con priorità maggiore
  - Periodicamente viene controllato lo stato di salute.
  - Un fallimento corrisponde a una riduzione di priorità.

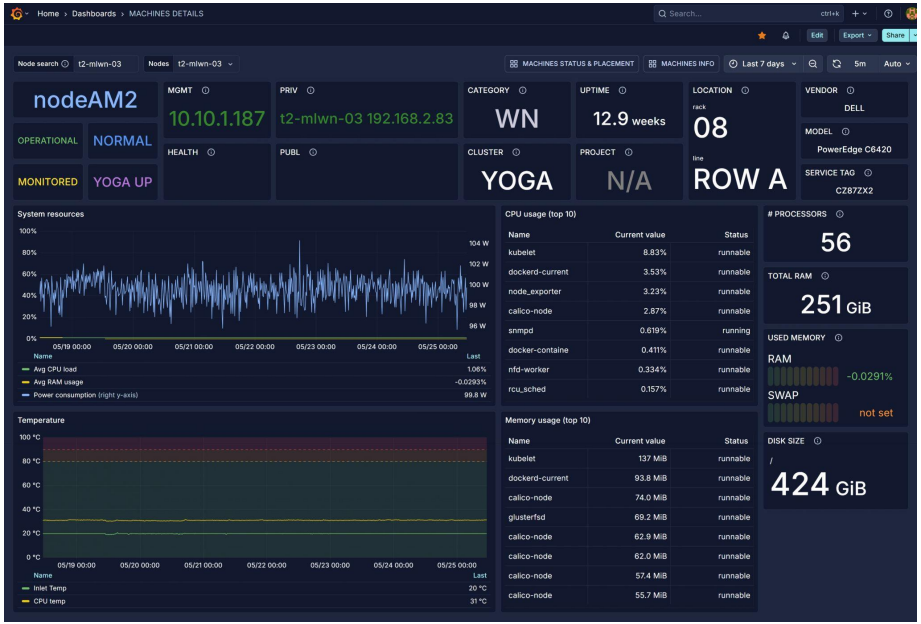


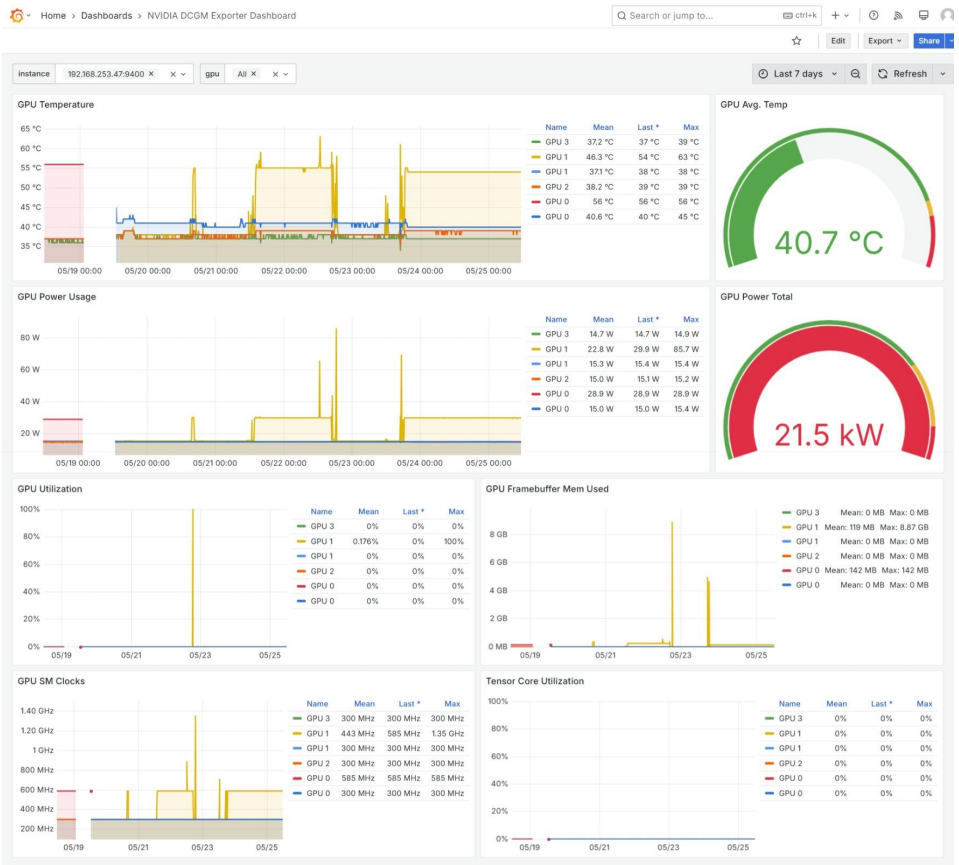
Stato dei lavori: GlusterFS

GlusterFS: è stato scelto, come primo test, un distributed volume per avere spazio aggregando bricks e rinunciando alla ridondanza



Monitoring





Ogni nodo viene monitorato dal punto di vista dello stato e dal punto di vista dell'utilizzo delle risorse

ToDo

- **3 control planes in High Availability**
- **/home directory su GlusterFS**
- **Storage dedicato ai dati**
- **Deployment dei nodi**
- **Autenticazione via ET-IAM**
- **Monitoring**

Desiderata:

- Installazione di Jupyter Notebook (primo access model che renderemo disponibile)
- Installazione di Rucio client
- Installazione di VRE
- …

The work is in progress...

... stay tuned

~ Thank you for the attention ~