

# WP1 Quick [re]start

ML\_INFNO, WP1

06/02/2023

# Risorse/infrastrutture

---

## Nuovo Hardware:

- 1 server (1TB, 128 core), 3 x A100, 5xFPGA (Alveo U50)
  - Ancora non arriva il Server da E4.
  - Quando arriva servirà probabilmente un down di un HV
    - Probabile redistribuzione dei devices tra i server

# Piani per il 2023

---

- **Cosa:**
  - superare alcuni problemi correnti:
    - Allocazione statica delle risorse allo end user (elevati idle time)
    - Vista / reporting delle allocazioni (Chi ha in uso cosa e da quando?)
- **Come:**
  - End User accede a container (ssh) o Jupyter (web).
  - Risorse istanziate via K8s
  - Integrazione con Batch System (Alternanza uso interattivo / batch )
- **Serve:**
  - Persistenza: salvare lo stato di un container su storage condiviso, in modo che possa ripartire successivamente e/o "altrove"
  - Rivedere / complementare Raccolta metriche di monitoring / accounting
  - Integrazione con batch system
  - Policy di alternanza Interattivo / batch
- **Per arrivare a:**
  - JupyterLab as a Service with platform-level persistency
  - Opportunistic batch resources

# Prossimamente...

Use Case per CNAF: zombie jobs early detection

