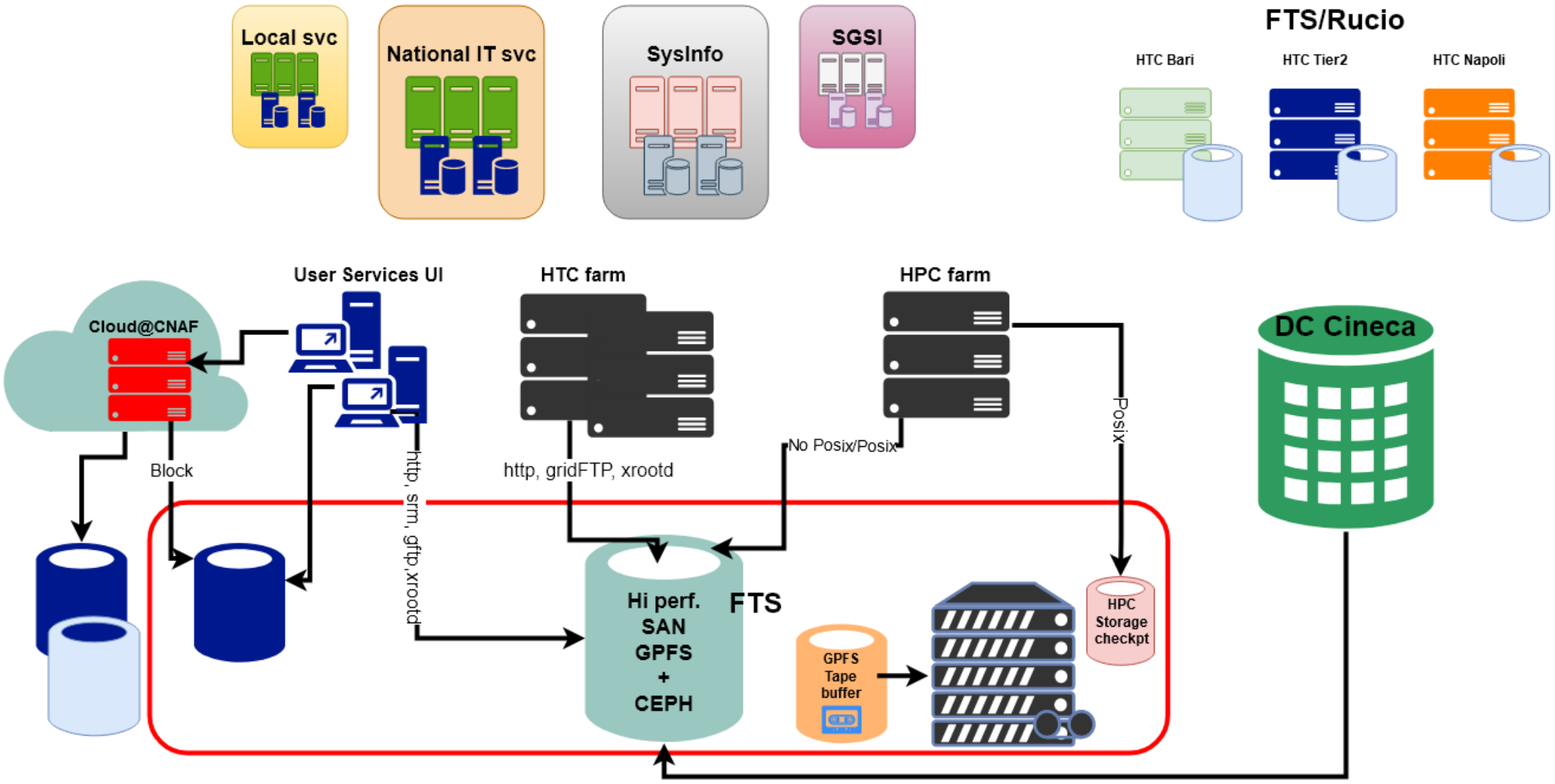
INFN CNAF
Istituto Nazionale di Fisica Nucleare
Centro Nazionale per la Ricerca e lo Sviluppo nelle Tecnologie Informatiche e Telematiche

Tecnopolo T2.1 – Offerta Tecnologica per il Computing

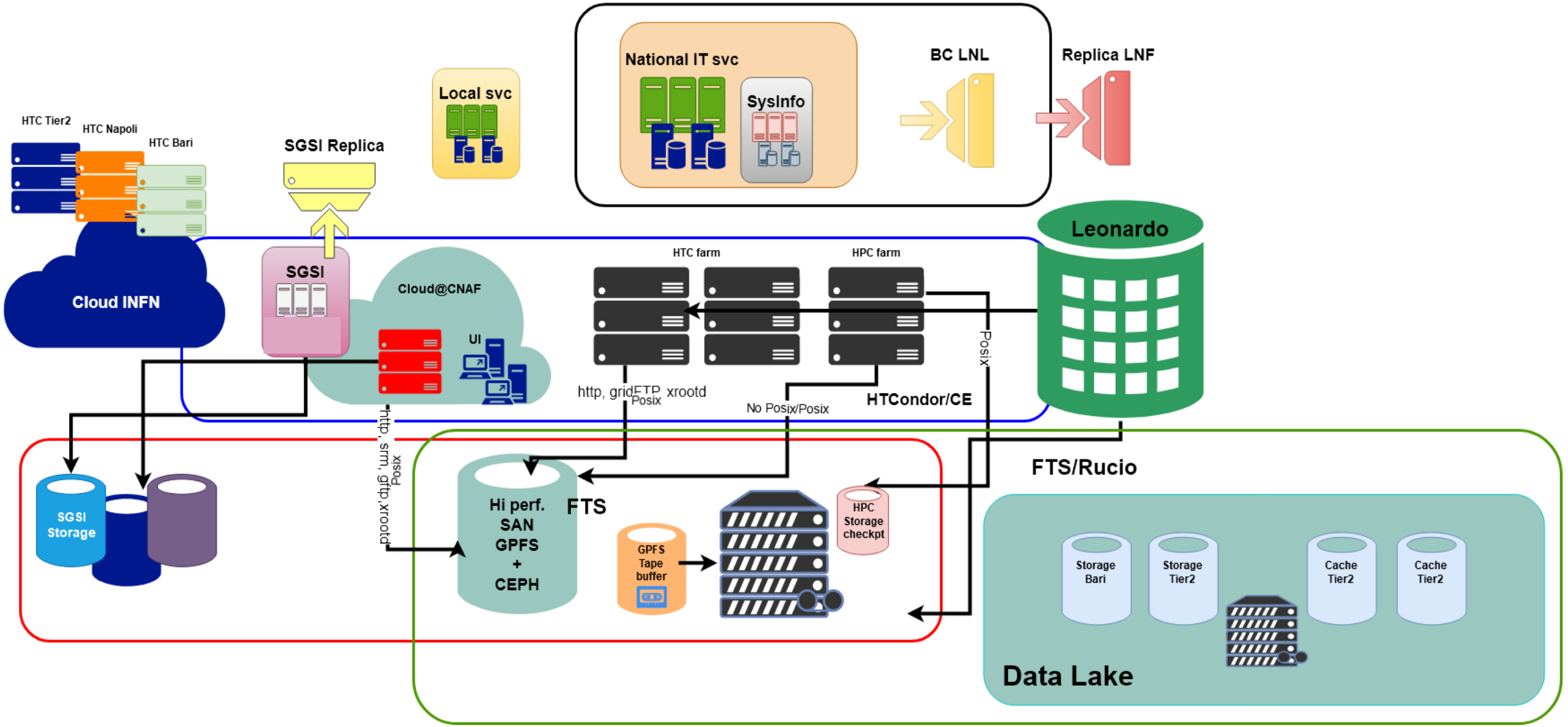
Istituto Nazionale di Fisica Nucleare
Centro Nazionale per la Ricerca e lo Sviluppo nelle Tecnologie Informatiche e Telematiche

Architettura di alto livello

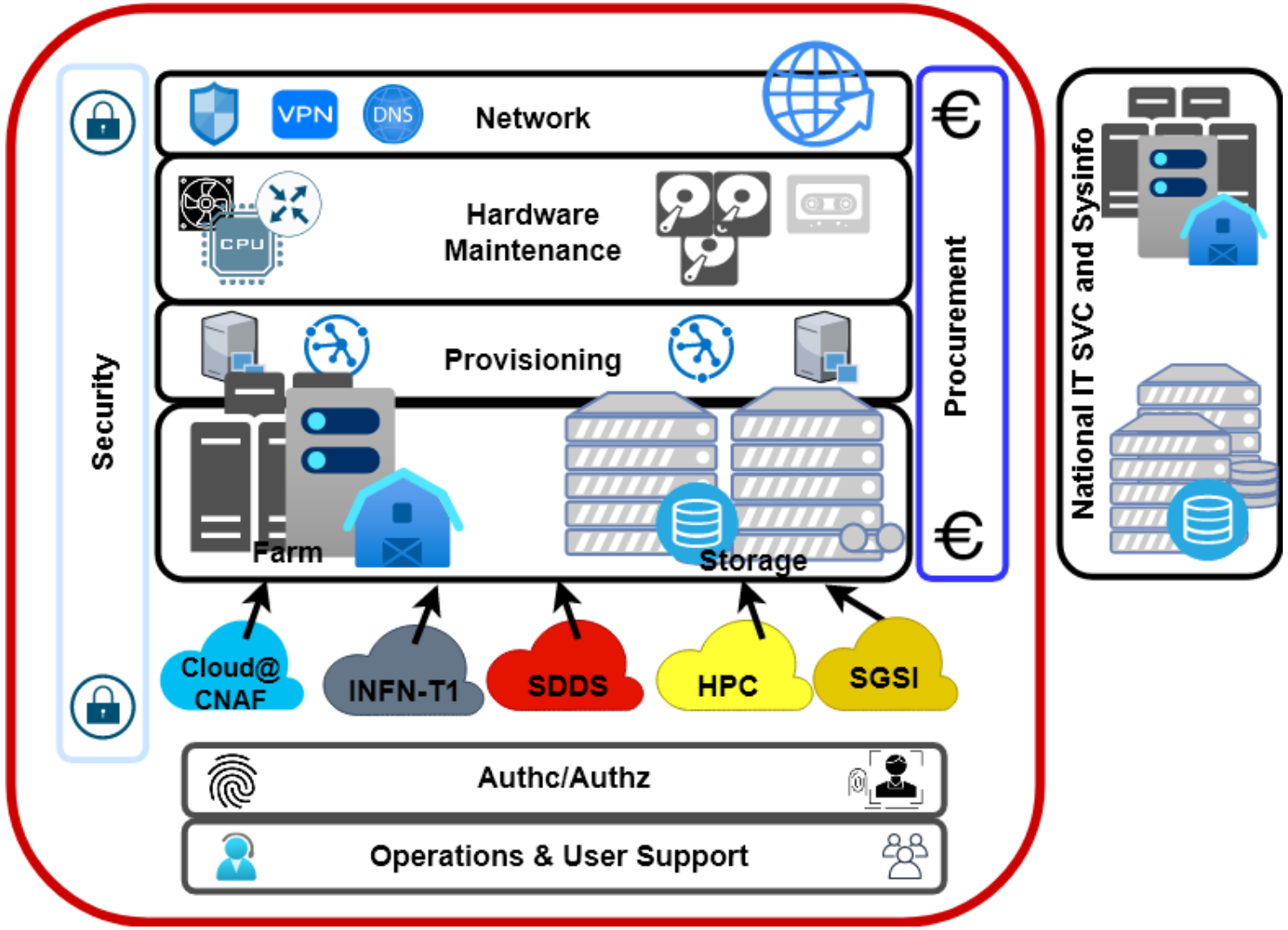
Architettura attuale



Architettura proposta



Schema logico-funzionale del nuovo DC



INFN CNAF
Istituto Nazionale di Fisica Nucleare
Centro Nazionale per la Ricerca e lo Sviluppo nelle Tecnologie Informatiche e Telematiche

Servizi per il calcolo nel nuovo datacenter

Servizi per il computing - 1

- Calcolo HTC
  - Via HTCondor/CondorCE
    - SLA e Computing model conformi a WLCG
      - Incluso accesso allo storage
    - Calcolo per comunità HEP, Astroparticle, GW et al.
      - Sottomissione via Grid e locale per VO con computing model mutuato da WLCG
    - Supporto a container
- Calcolo HPC
  - Sottomissione da CondorCE verso Leonardo (T2.4)
  - Sottomissione via Grid e in locale(SLURM) verso un cluster HPC-CNAF dotato di GPU e bassa latenza
    - Da integrare l'accesso alla SAN ad alte prestazioni, mantenendo un'area scratch per checkpoint
    - Use Case
      - VO che accedono alla farm HTC e che necessitano anche di GPU
      - Prove, test e sviluppo in locale prima di andare su supercalcolatori (use case dei cluster attuali)
  - Computing su GPU (Batch)

Servizi per il computing - 2

- Accesso Cloud IaaS via Cloud@CNAF
  - Omogeneizzazione dei vari cluster virtuali ad uso interno
  - Provisioning di User Interface
  - Richieste "spot" di analisi in interattivo
    - Le UI «carrozzate»
  - Servizi specifici e infrastrutturali per VO
    - DBs, dedicated monitoring, dedicated accounting, low latency
    - IAM, voms, «vobox» etc..
    - Cluster K8S con gestione centralizzata: Virgo
  - Servizi e testbed per progetti interni ed esterni:
    - INDIGO-DataCloud, XDC, DEEP, IoTwins, SmartChain, SUPER
    - EEE
  - Supporto attività R&D – progetti staff CNAF
  - Supporto corsi formazione

Servizi per il computing - 3

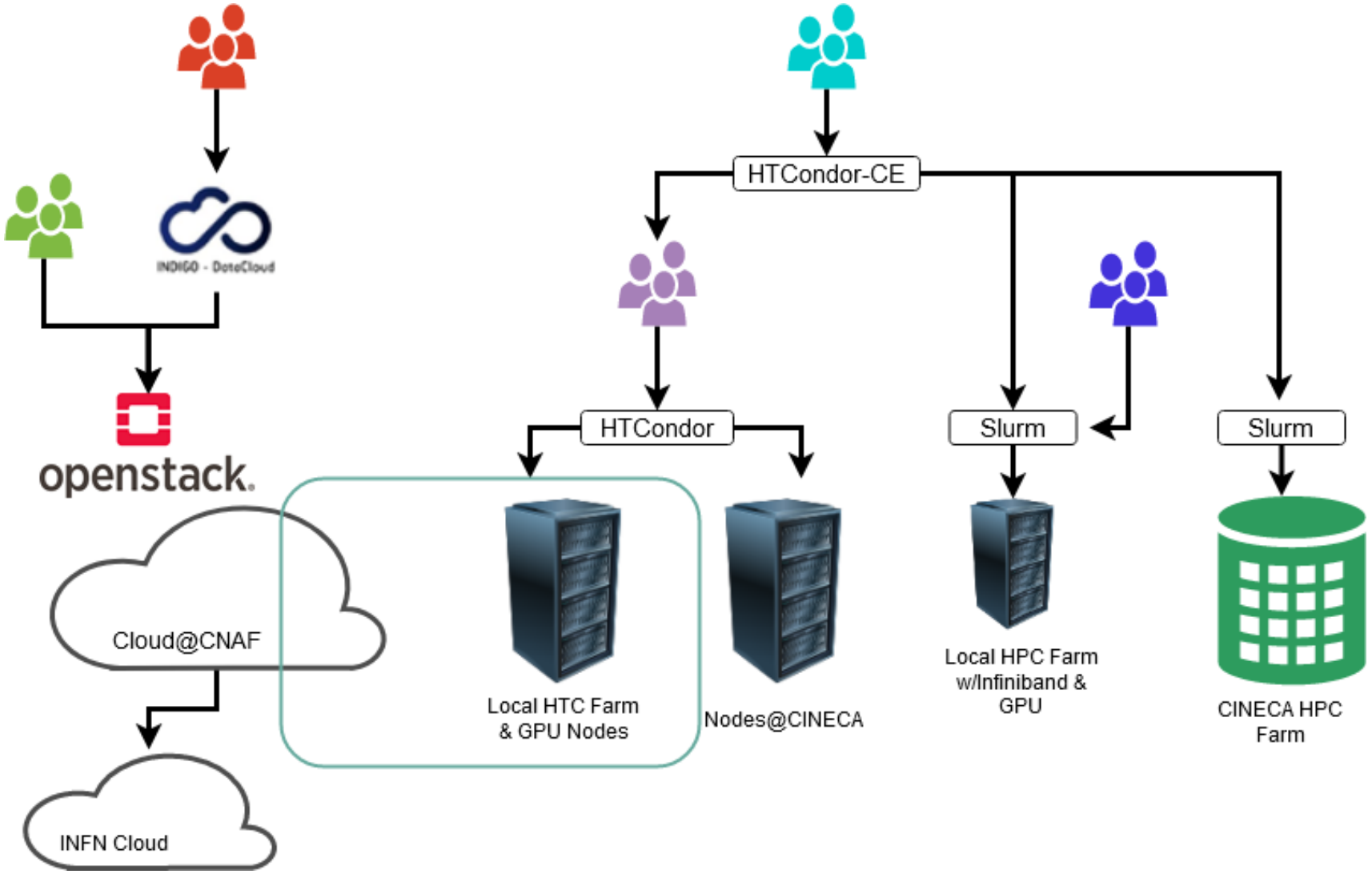
- Accesso Cloud PaaS via Dashboard (Dashboard INDIGO PaaS Orchestrator) con o senza federazione INFN-CLOUD
  - Deployment di cluster dedicati a piccole collaborazioni
    - i.e. dynamic cluster su K8s
    - Cluster per collaborazioni temporanee (progetti)
  - Risorse per low Latency analysis
  - Espansione dinamica di farm HTC (anche remote)
  - Risorse CPU/GPU dedicate per didattica tecniche ML/DL (ML_INFN)

- Provisioning of Platforms for Big Data Analytics
  - Monitoring del datacenter
    - Predictive maintenance

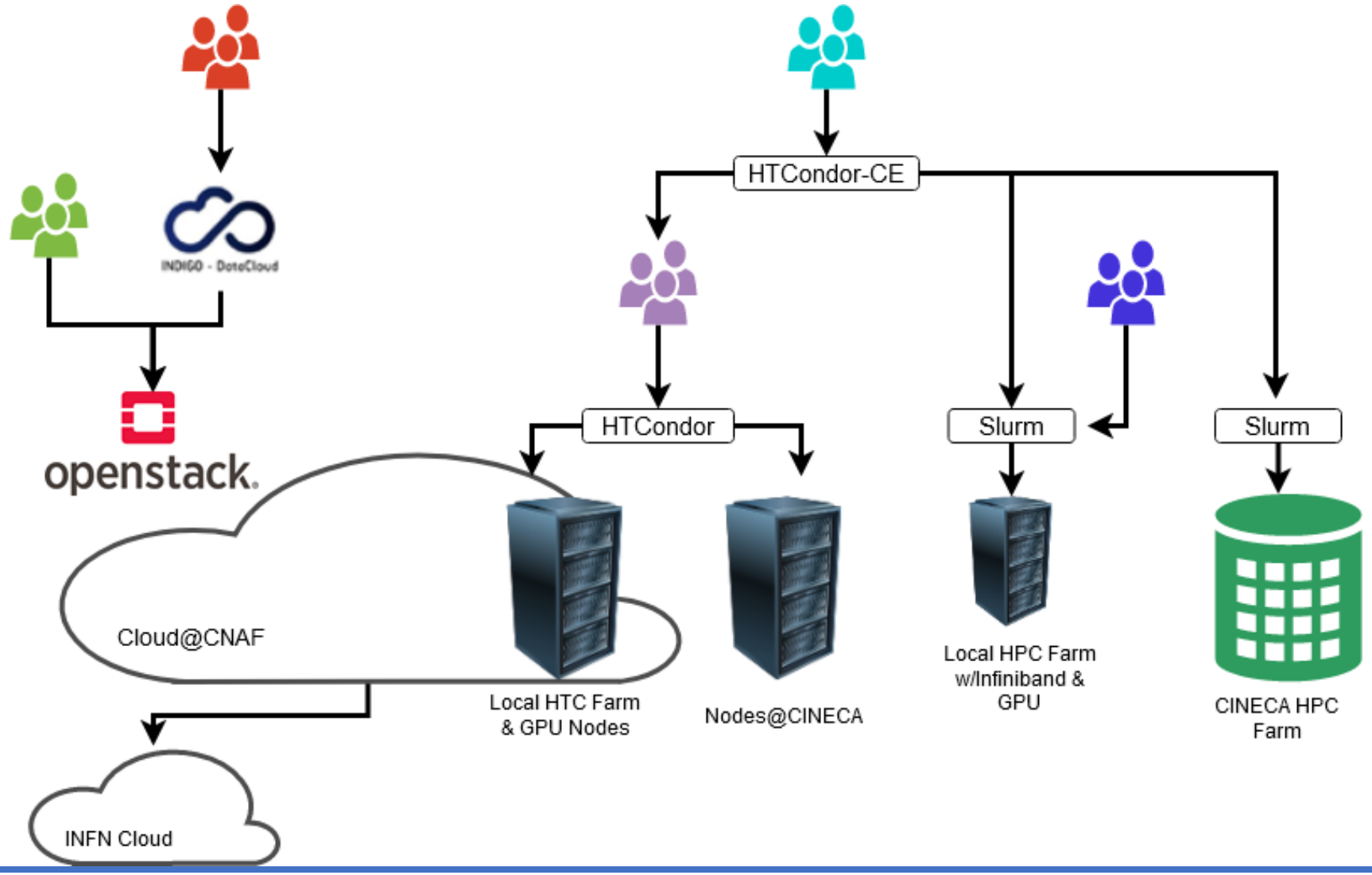
- HPC in the Cloud?
  - Fino ad ora esperienza limita su singoli nodi con GPU
    - Piccolo cluster IoTwins con InfiniBand

- Accesso risorse nell'area sicura “SGSI” (vedi presentazione dedicata)

Accesso al computing



Accesso al computing



Istituto Nazionale di Fisica Nucleare
Centro Nazionale per la Ricerca e lo Sviluppo nelle Tecnologie Informatiche e Telematiche

Possibili Soluzioni tecnologiche

Batch+Cloud

- Come ripartire/unificare le risorse tra farm HTC e risorse Cloud?
  - Soluzioni sotto investigazione
    - Virtualizzazione di tutta la farm su Cloud
      - Ironic per WN bare metal
    - WN HTCondor come pod di k8s su bare metal
    - WN HTCondor come pod di k8s instanziato su VM OpenStack
- Questioni da dirimere
  - Accesso risorse con GPU
  - Come segregare le reti dedicate LHCONE/LHCOPN?
  - Come garantire la banda verso SAN da dentro al cloud?
    - Configurazione Neutron nodes ?
  - Federazione con INFN-CLOUD
  - Auto-scaling dell'infrastruttura

Evoluzione calcolo batch

- **Obiettivo**: provisioning attraverso orchestratori
  - evitando overhead di virtualizzazione
- **Approccio**: WN istanziati come container su bare metal
  - Ridurre al minimo la perdita di prestazioni rispetto agli attuali WN del CNAF
- **Strumenti** considerati:
  - Ironic:Openstack, per provisioning a livello bare-metal
  - Docker: come WN standard del T1
  - K8S: in seguito, ri-organizzare WN docker "monolitico" in POD di k8s

Evoluzione calcolo batch - 2

- **Metodo**: prendere a modello esperienze già presenti nella comunità (es. DODAS)
- **Migliorie possibili**
  - WN container "parziali": General Purpose ma confinati a una porzione del bare metal (es. ¼ delle risorse del nodo Host)
  - WN container "dedicati a User Groups": specifici per job di solo alcune VO
  - WN container "dedicati a WorkLoads": specifici per determinate attività (es. Accesso a GPU; IO intensive, CPU bound, RAM bound)
- **Aspetti da considerare**
  - Orchestratori di container/POD possono/devono fornire la conoscenza dei nodi su cui istanziare
    - Sovrapposizione parziale con modello HTCondor: esistono due livelli di amministrazione del cluster, in concomitanza
  - Es: aggiornamento kernel/immagini (del BM e/o del container) la gestione passa attraverso diversi livelli di gestione

Tempistiche e risorse umane

Le problematiche sono tante e interdipendenti, milestone e stima effort in via di definizione…

- Ironic
  - Testbed al CNAF a settembre 2021, 1.5 FTE (farming, sdds)
- Unificazione farm HTC/HPC
  - 2 mesi a trasferimento avvenuto, 1 FTE (farming, us)
- Espansione dinamica
  - Già possibile, 0.5 FTE per implementare tutti gli scenari
- Orchestrazione – K8s integration
  - Testbed al CNAF ad aprile 2021, 1.5 FTE (farming, sdds)
- Valutazione performance di accesso allo storage dalla nuova infrstruttura (farm t1 in particolare)