The new INFN Data Center at Bologna Tecnopolo

L.dell'Agnello, T.Boccali, A.Cherici, S.Zani, L.Scarponi, V.Sapunenko, D.Cesini
On behalf of the CNAF-Reloaded team

INFN-CCR Workshop
24 May 2022 - Paestum

A brand new CNAF datacenter

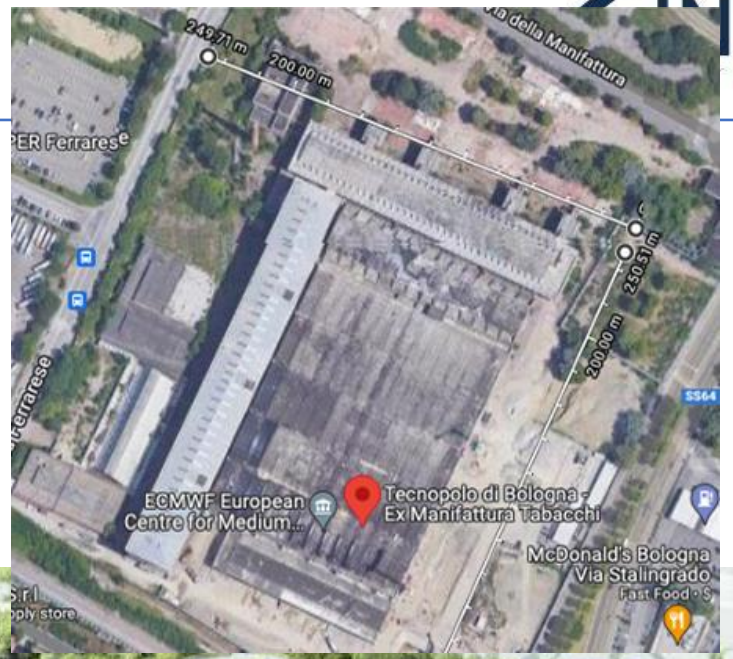
- Renew infrastructures to be ready for the HL-LHC era
  - up to ~ 2035 and beyond
- Use more compact computing
  - from today's ~ 20 kW/rack to 80 or more
  - Integration with CINECA-Leonardo
- Lower the PUE (*power usage effectiveness*)
  - be greener
  - Targeting 1.08-1.10
- Extend and expand networking for a future-proof infrastructure

The opportunities ….

- In **2017**, Bologna won a bid to host the "*European Centre for Medium-Range Weather Forecasts*"

- The Emilia Romagna region decided to repurpose the "*Manifattura Tabacchi*" area to host a technology district, hosting ECMWF and more

Roughly 250x250 m²

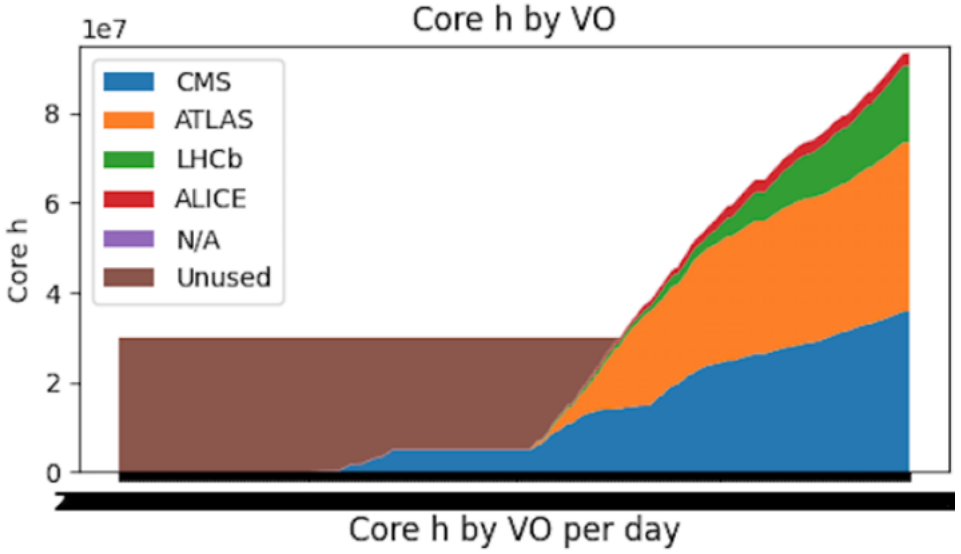


How it will be

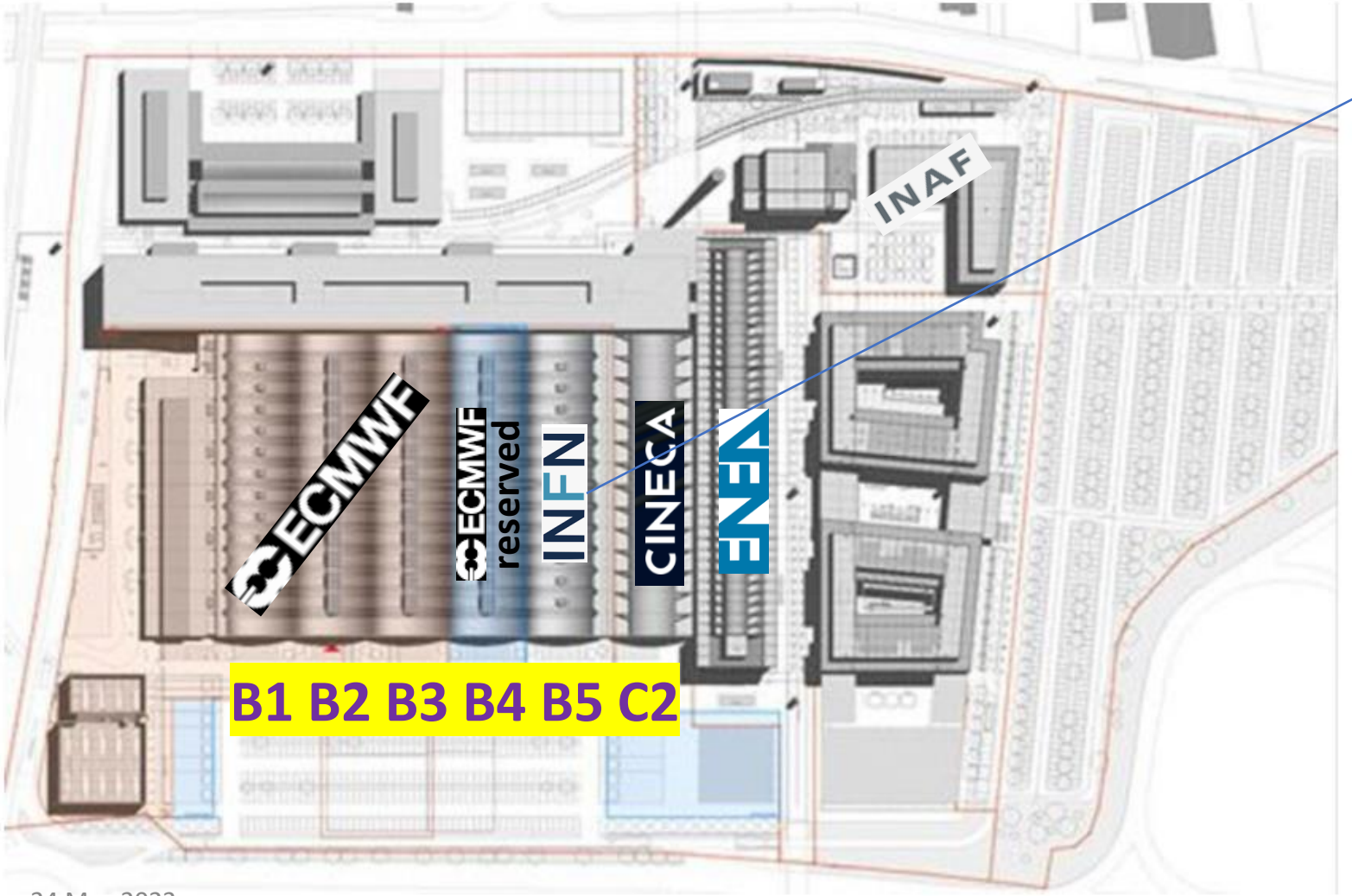
The opportunities…

- **Summer 2019**: Italy (CINECA, INFN, SISSA) won a bid for one of the three EuroHPC JU pre-exascale system
- Leonardo (250 PFLOPS), to be deployed at the Tecnopolo in the next months
  - Intel CPUs (good for us!) and Nvidia GPU
    - ~ 7000 last generation Xeon CPUs (~10 MHS06)
    - ~14000 Nvidia A200
    - Tight integration with CNAF storage (~2 Tbps)
- CINECA/INFN have a long collaboration history
  - A fraction of the current CPU power of CNAF is indeed hosted by CINECA

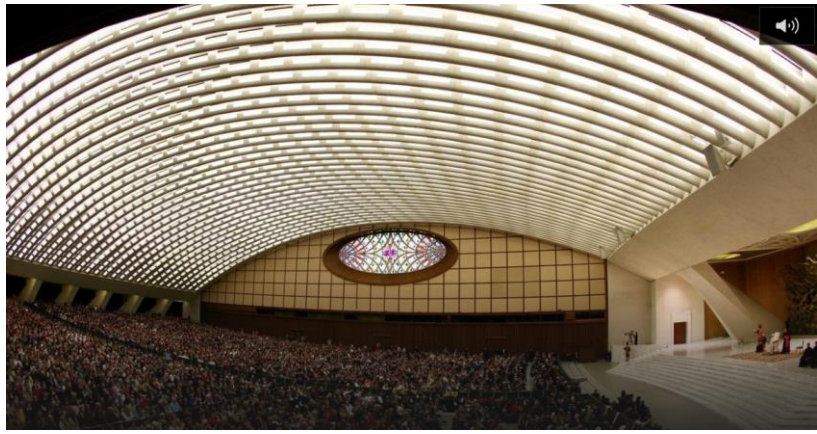
- CINECA and INFN already collaborate in the direction of enabling HPC systems for HEP
  - 1 PRACE grant (2019) demonstrated the feasibility of transparent LHC computing on the Marconi system with more than 90 Million Core Hours delivered
  - Marconi100 is the first IBM/Power system validated for physics utilization at LHC (with CMS)



What can the Tecnopolo host?



Each of the 6 “botti” (barrels) is ~5000m^2 of usable IT space

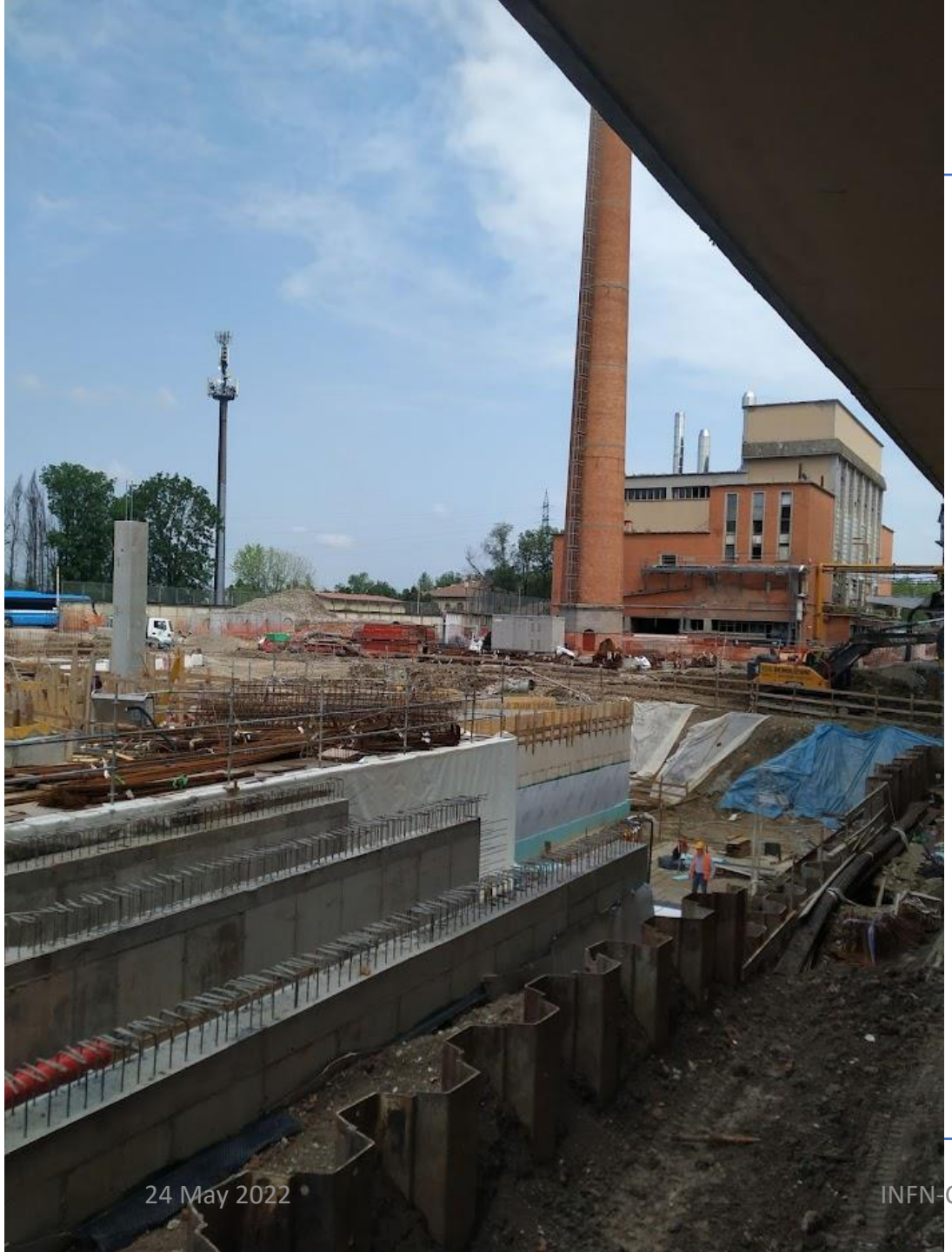


Same architect and design of the ”Sala Nervi” in the Vatican

The INFN+CINECA project

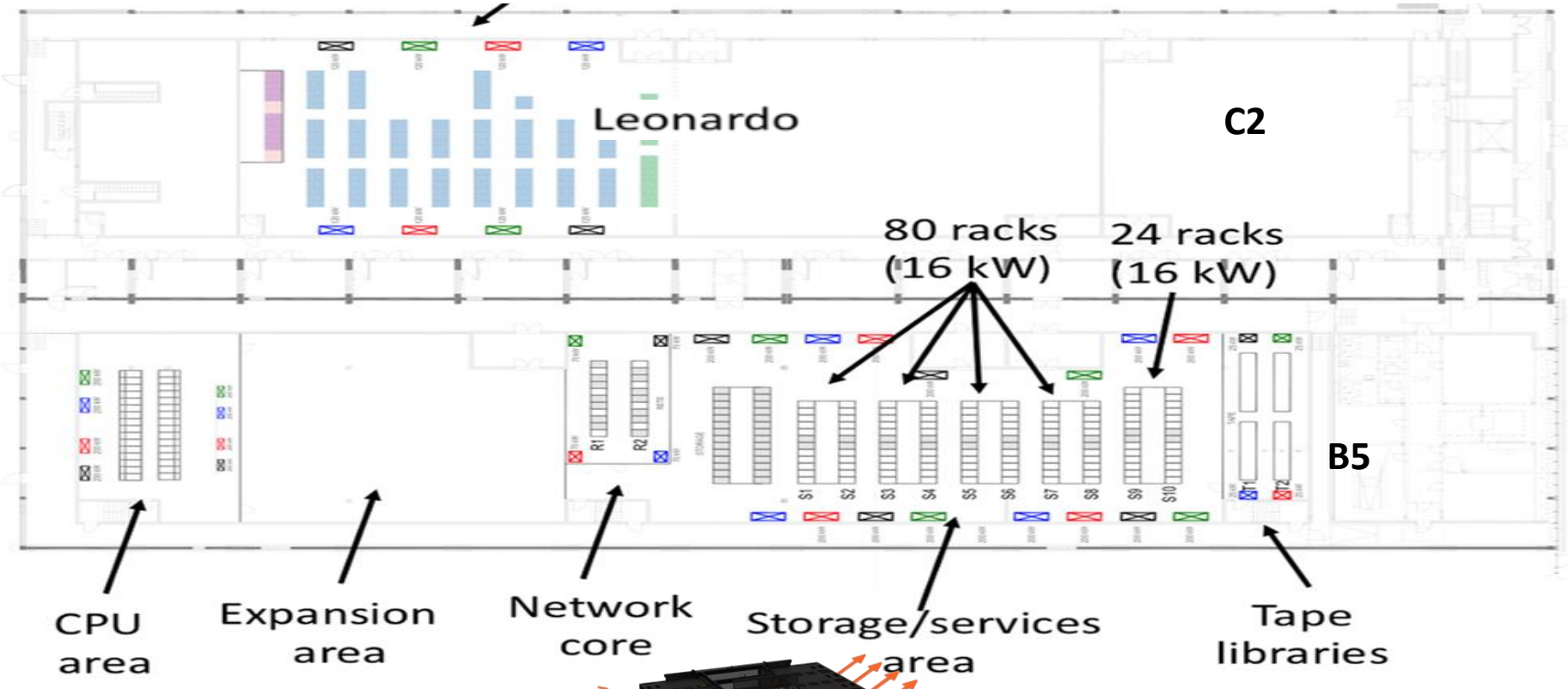
- The CINECA (“C2”) and INFN (“B5”) barrels are expected to be ready by
  - ~mid 2022 (CINECA)
  - ~First half 2023 (INFN)
- Two phases expected
  - **Phase-1 (2023-2025)**: Leonardo + T1-CNAF relocated. Total 10+3MW
  - **Phase-2 (2025+)**: infrastructure up to 24 MW (10MW INFN) ready for post-exascale and for HL_LHC


edilcentro



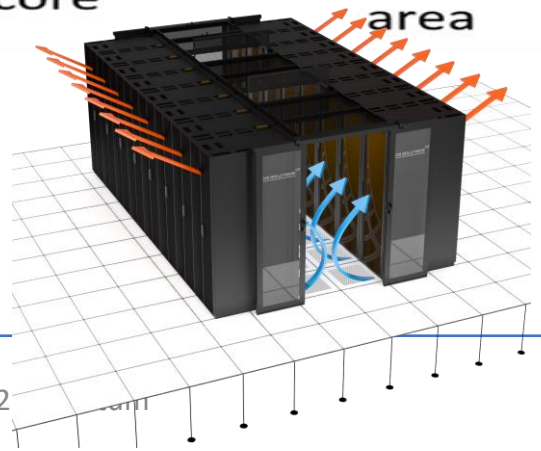
CINECA AND INFN (C2 AND B5)

- Even very pessimistic (unrealistic) HL-LHC extrapolations to 3x over flat budget would fit in B5, using the expansion area
- The CPU area will be initially unused
- But can later host up to 3MW of CPUs via 42 DLC high density racks
  - **And more in the expansion areas, if needed**



DLC 80kW

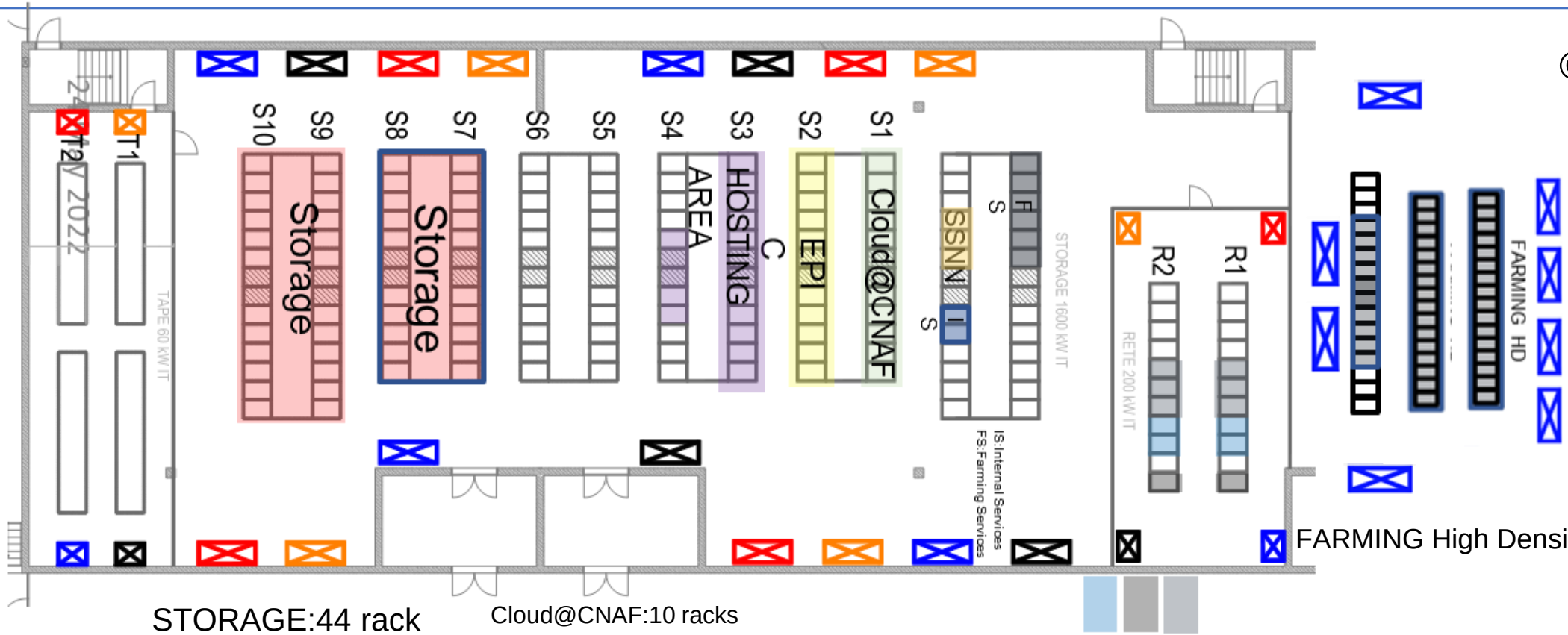
H2O: 40-50°C



H2O: 19-26°C

Total of 20 CRAH 200 kW each
Low density: 2MW 3+1
Cold corridor:25°C
Env: 32°C

Data Center Layout



© S.Zani

FARMING High Density (42 racks)

STORAGE:44 rack

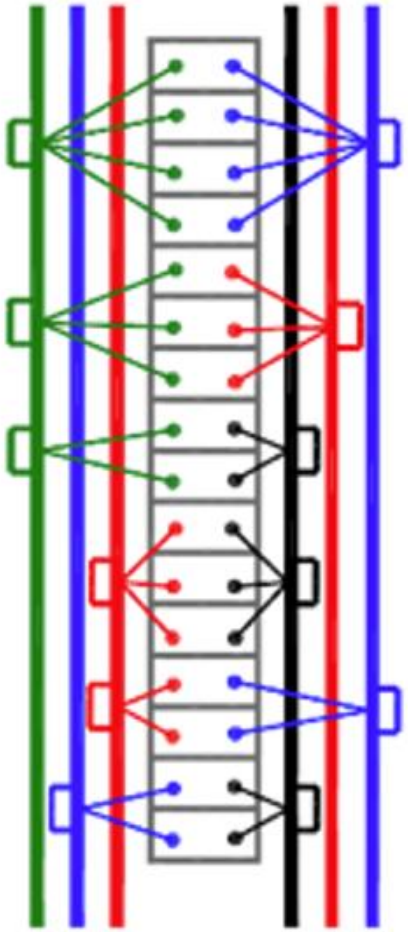
TAPE area

Cloud@CNAF:10 racks
EPIC:5 rack (up to 10)
SSNN+Internal Services:5 racks
Farming Services:6 racks
Hosting: 13 racks

Network CORE ZONE
Internal Passive Cabling: 6 rack (3+3)
Patch Panel: 2 rack
Switches and Routers Fase 1: 4 rack
2 Network Racks per Computing ROW

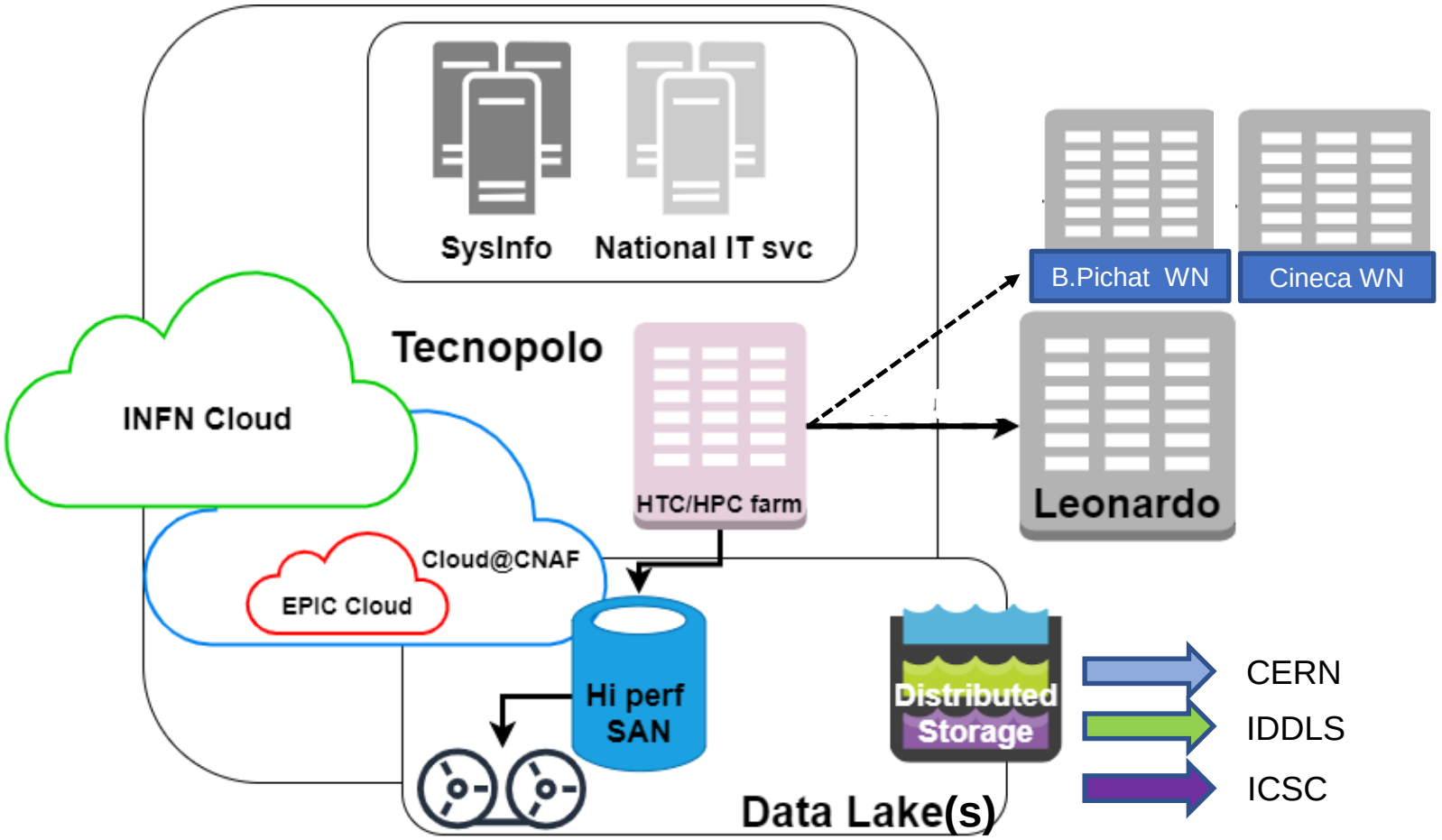
Low density: 156 rack di cui 132 per risorse IT, e 24 per apparecchiature di rete - 42U P:1200mm – L:750-800mm

Electrical distribution



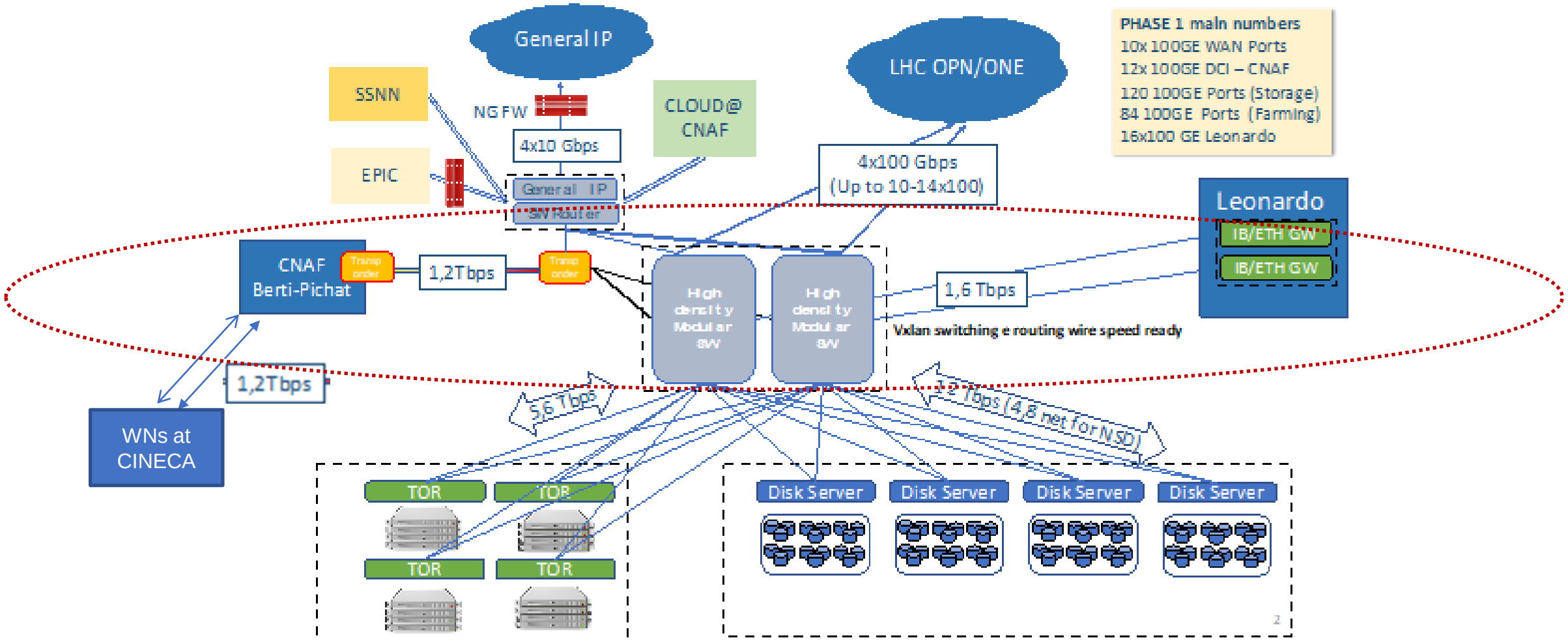
- Phase 1: Hall B5 - 3 MW
  - under UPS and diesel generator
  - distribution with 3 + 1 redundancy
    - using 4 UPS
      - 1 MW each
  - Low Density: 24 powerline - 1000A each
  - High Density: 6 powerline - 1000A each
  - Phase 2 (2025+) ➔ 10MW

A “distributed” datacenter



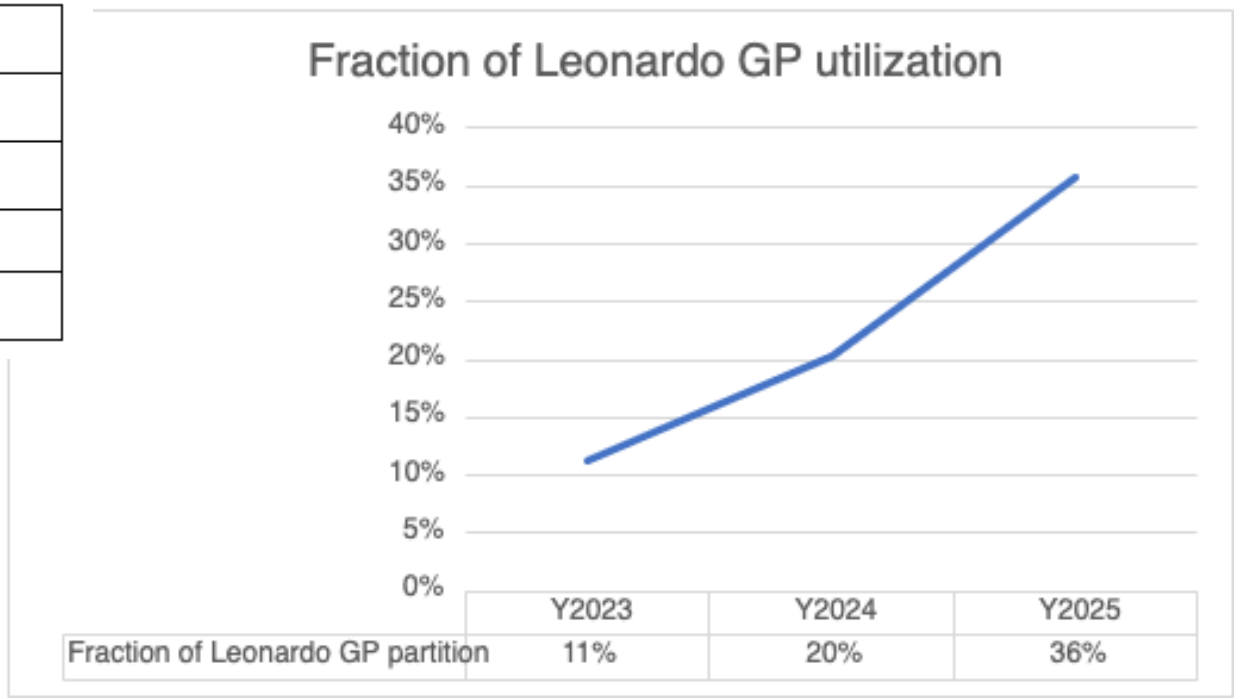
- Multiple “heads”
  - CNAF Tecnopole
  - Cineca Leonardo
  - INFN-CLOUD
  - Data-lake(s)
    - DCI with INFN sites (IDDLS)
    - DCI with CERN
    - ICSC (centro nazionale)?
  - Legacy CNAF during migration in the initial stage
    - B.Pichat CNAF
    - CNAF WN@CINECA
  - Business continuity
    - EPIC
    - National Services

Network layout



Growth profile of installed resources

YEAR	CPU	DISK	TAPE
	kHS06	PB-N	PB
2023	820	78	172
2024	990	94	206
2025	1320	110	247



- To avoid new CPU acquisition and using Leonardo to fullfill the pledges 2023-2025:
  - 2023 - 11% (165 nodes) of Leonardo GP
  - 2024 - 20% (300 nodes)
  - 2025 - 36% (540 nodes)
- CINECA agreed for 2023, 2024
  - Not possible for 2025
    - We will exploit the farm HD area, DLC CPU

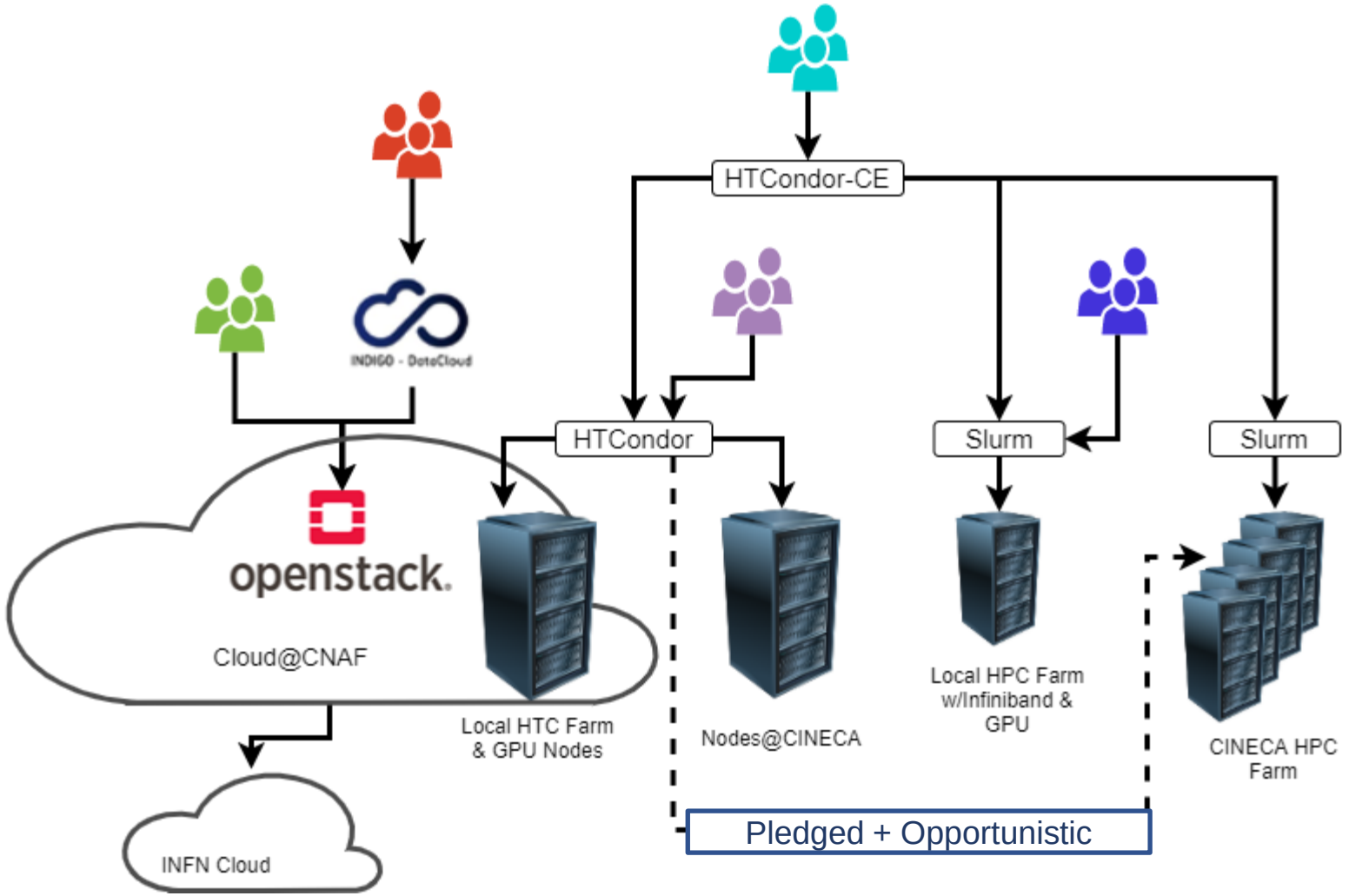
11% = 165 nodes Leonardo GP (~ 2 rack)
20% = 300 nodes Leonardo GP (~ 4 rack)

How to connect Leonardo

- Open issue:
  - Leonardo nodes connected via IB
  - CNAF connected via ETH
- The plan is to use 2 IB-Ethernet gateways (Mellanox Skyway)
  - aggregated bandwidth 1.6 Tbps
  - Testing at scale needed

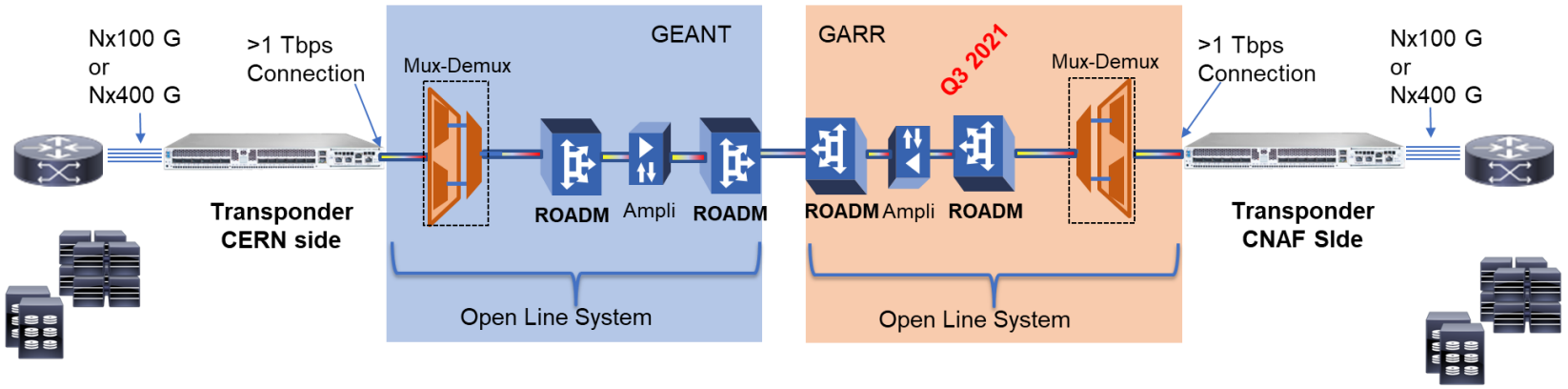
- The skyways
  - 8 connectx-6VPI dual head NICs
    - One IB, one Ethernet
  - Nvidia guarantees 8x100Gbps on the eth side
    - Could reach 200Gbps but currently there are limitations on the PCI bus
  - Can be configured as L3 switches with 8 ports at 100Gbs in LACP load sharing
  - VLAN tagging is not possible
    - as other functions typical of Ethernet switches
  - Can map an IB fabric to 1 IP subnet
  - Max MTU on the Eth side is 4092
    - GPFS interoperability?!?
  - Does not support IPv6
    - But should be fixed in a near future
  - 2 Skyways can be configured as 2 distinct switches or as redundant 16 ethernet ports system

Farm + Cloud integration



- Few access points for several different infrastructures
  - Cloud@CNAF
  - Cloud@INFN
  - Local HTC farm
  - Local HPC farms
  - CNAF WN@CINECA
  - Leonardo partitions
- Based on
  - HTCondor
  - SLURM
  - HTC-SLURM connectors
  - INDIGO PaaS + Openstack

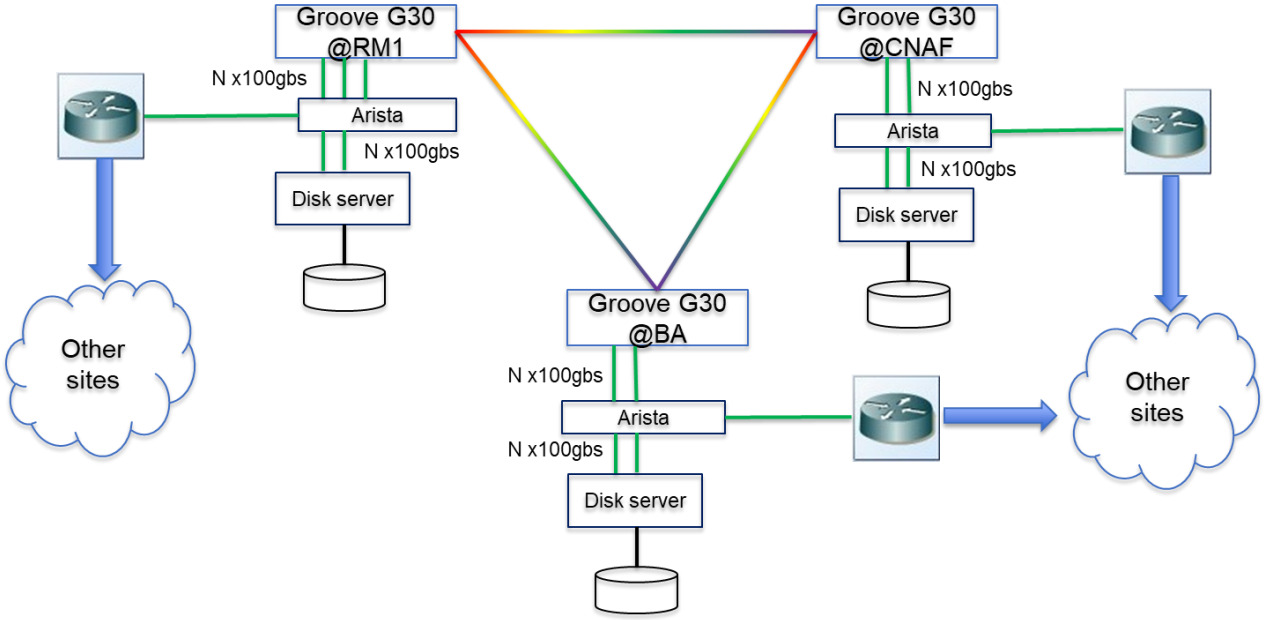
DCIs - Data Center Interconnect



DCI CERN-CNAF

R&D proposal to implement a Terabit DCI between CERN and CNAF

Packet/optical **transponder inside the datacenters interconnected by a Open Line System** transport provided by GARR+GEANT



IDDLS (Italian Distributed Data Lake for Science) Architecture

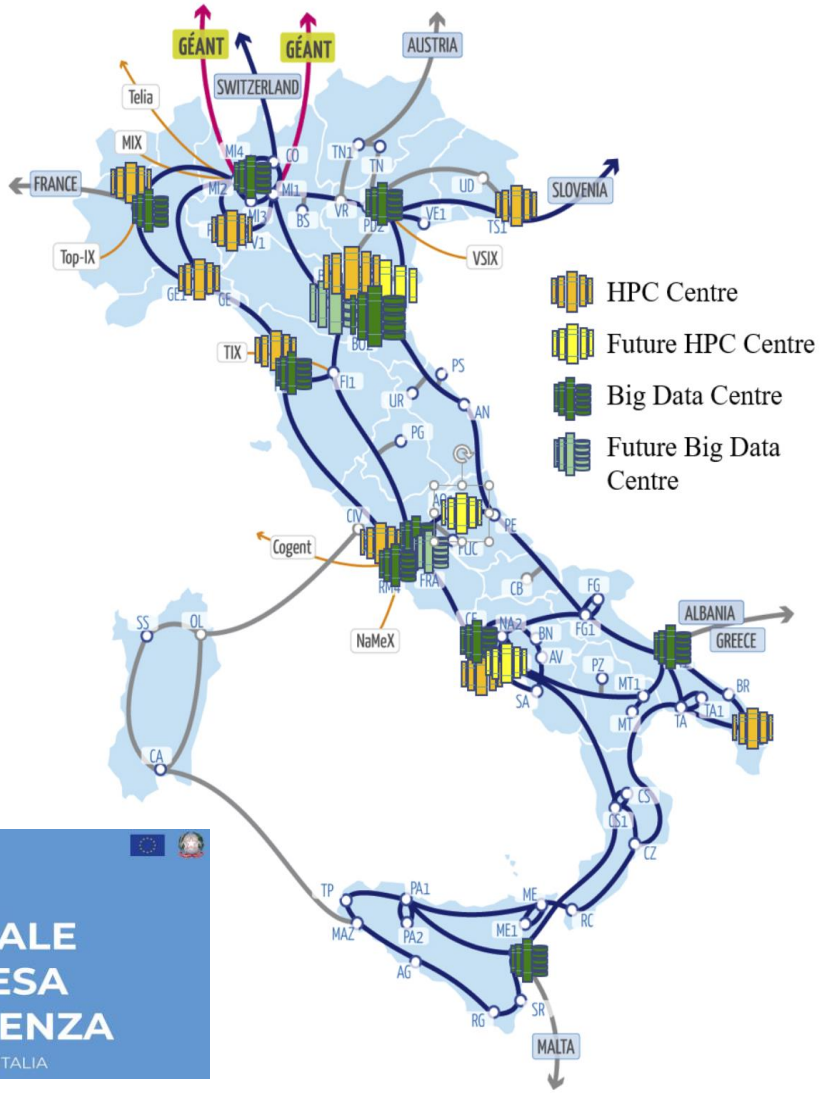
INFN-GARR collaboration to realize a prototype of an Italian Datalake exploiting

Live Migration

- CPU
  - Transparent to users
    - Maintain WNs@CINECA (420kHS06)
    - Move legacy from current site to Tecnopolo (120 kHS06)
    - Start using Leonardo
- Disk
  - Install 30PB at Tecnopolo
  - Move data on the new storage
  - Move legacy hw to Tecnopolo (50-55PB)
- TAPE
  - Move one library at a time
    - Cannot read from the library that is being moved
    - Write on buffer will be transparent

ICSC and the Tecnopolo

- Headquarters at the Tecnopolo
- 34 Research institutions and Universities
- 15 major Italian companies
- Over 1500 researchers committed
- Deployment of a production-level **Quantum system** (emulator or real hardware, tbd)
- Deploy and operate the resources via a national level **Datalake**



Companies

LEONARDO · INTESA SANPAOLO · FERROVIE DELLO STATO ITALIANE · fondazione innovazione urbana · eni · FINCANTIERI · autostrade per l'italia · IFAB · UnipolSai ASSICURAZIONI · HUMANITAS RESEARCH HOSPITAL · ENGINEERING · ThalesAlenia Space · sogei · Terna Driving Energy · UPMC LIFE CHANGING MEDICINE

National Institutes

INFN · CINECA · ENEA · Consiglio Nazionale delle Ricerche · Consortium GARR · INAF · INGV

PIANO NAZIONALE DI RIPRESA E RESILIENZA
#NEXTGENERATIONITALIA

Conclusion

- CNAF is moving to a new location
  - future-proof due to high expansion possibilities
  - ready for the requirements of the next generation experiments
    - not only LHC and not only (astro) physics
- One datacenter composed of distributed facilities
  - HTC, HPC, Cloud,
  - datalakes and ICSC infrastructures
  - Few user entrypoints
- Closer integration with the CINECA HPC facilities

Data Valley: il video racconto di Stefano Accorsi, entrato nel Tecnopolo

https://www.youtube.com/watch?v=96TfXHCWxf8