

T1 highlights @ Meeting annuale ATLAS Italia Computing

28/11/2023

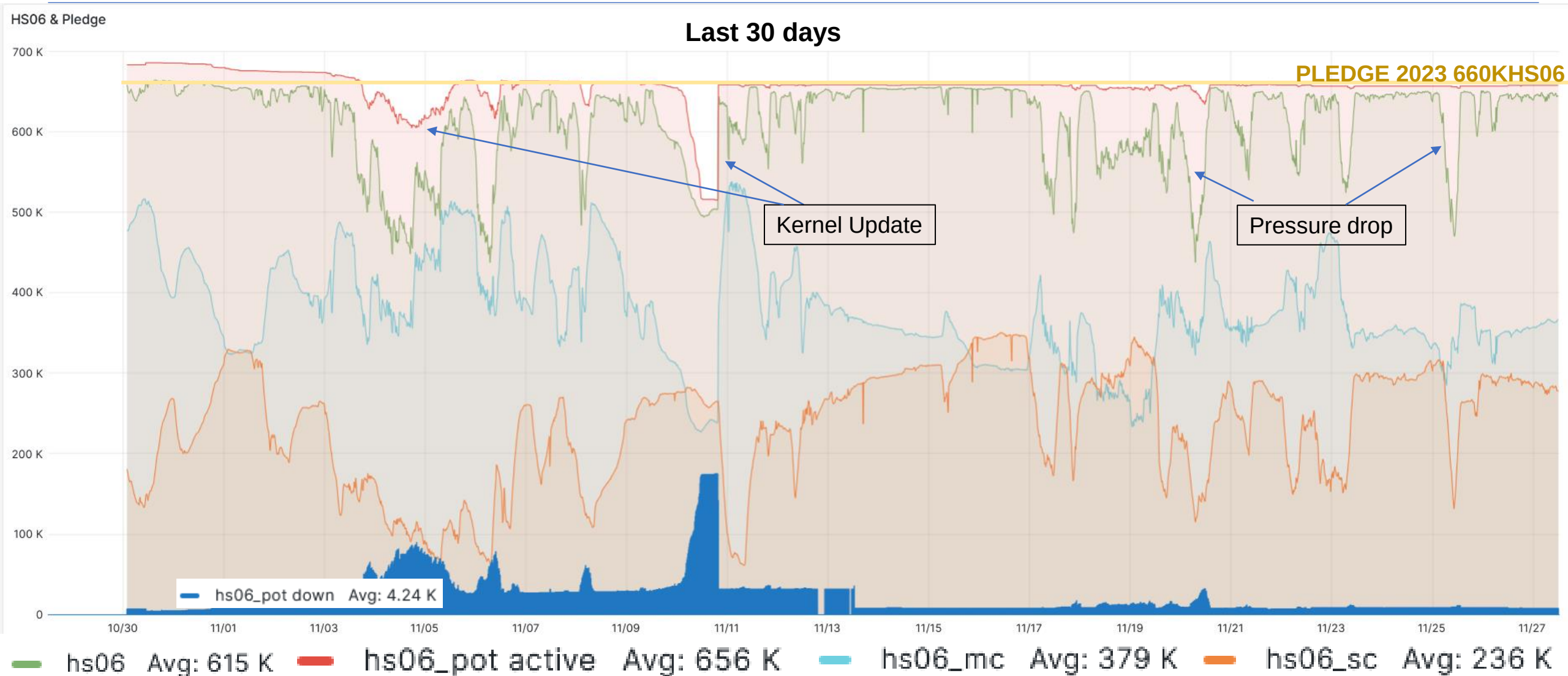
D.Cesini – INFN-CNAF

Resources@T1 2023-2024

ALL VO No Cloud	2023	2024	Delta
Pledge CPU (HS06)	660000	792000	132000
Pledge disk (TBN)	69576	82949	13373
Pledge tape (TB)	158282	193581	35299

ATLAS	2023	2024	Delta
Pledge CPU (HS06)	128700	136440	7740
Pledge disco (TBN)	12240	14670	2430
Pledge tape (TB)	31770	40680	8910

CPU Usage @T1 ALL VOs – no cloud



CPU in 2024 – Leonardo integration

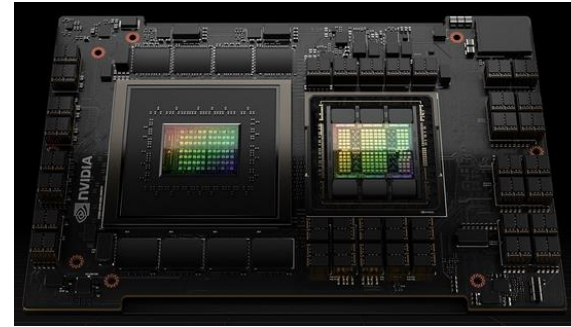
- No CPU acquisition in 2024
- We will use up to 300 Leonardo-GP@CINECA nodes
 - Dual 56 cores sockets Intel Sapphire Rapids
 - 112 cores/node
 - (16 x 32) GB DDR5 4800 MHz
 - 1680 Gflops/node (peak)
 - HS06 ??
 - >2000HS06/node
 - NVIDIA Mellanox HDR DragonFly++ 200Gb/s
 - No Ethernet
- Integration Plan
 - “inifnite” SLURM jobs to launch VM containing “our” Condor WN
- PCI pass-throught to see the IB cards on Leonardo
- Mellanox Skyway IB-ETH bridges to reach our LAN
 - 16 x 100Gbs



- › Standard 2U appliance
- › 1.6Tb/s solution
- › 8-port HDR/HDR100/EDR InfiniBand
- › 8-port 200/100Gb/s Ethernet

- 1 ARM node in production
 - 256 cores
 - 592 GB RAM
 - Soon will be upgraded to 1TB
 - 2 TB disk
 - **CMS**
 - Testing phase ended
 - In production since 23/11
 - **ATLAS**
 - Under testing
 - singulatrix el9 image under finalization on the ATLAS side
- 2 new ARM nodes in preparation
 - By the end of december
 - 256 cores
 - 1 TB RAM
 - 2 TB disk

- Coming soon: NVIDIA Grace Hopper Superchip



Key Features

- > 72-core NVIDIA Grace CPU
- > NVIDIA H100 Tensor Core GPU
- > Up to 480GB of LPDDR5X memory with error-correction code (ECC)
- > Supports 96GB of HBM3 or 144GB of HBM3e
- > Up to 624GB of fast-access memory
- > NVLink-C2C: 900GB/s of coherent memory

- Coming soon: NVIDIA Grace Superchip

Key Features

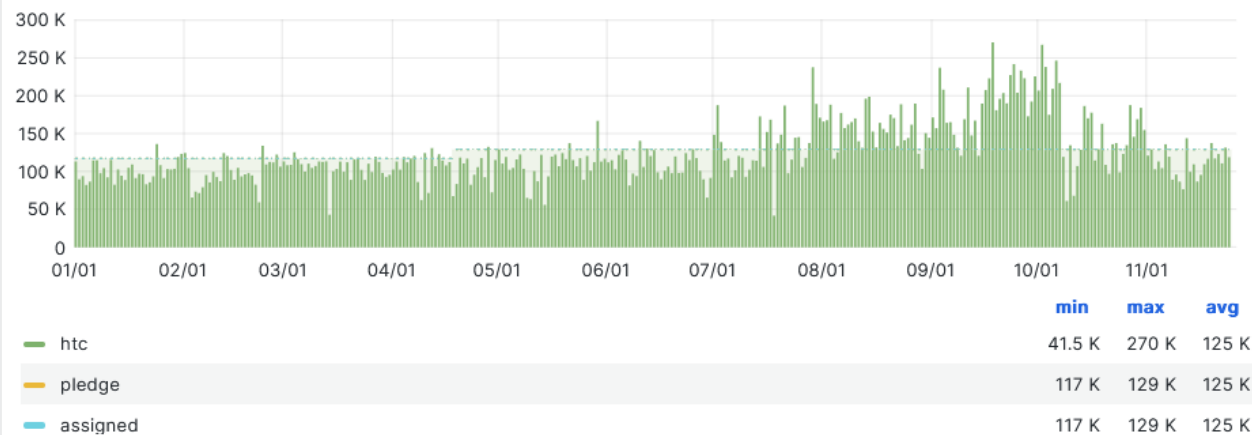
- > Up to 144 high-performance Arm Neoverse V2 Cores with 4x128b SVE2
- > High-performance NVIDIA Scalable Coherency Fabric with 3.2 terabytes per second (TB/s) bisection bandwidth

- > Up to 960 gigabytes (GB) of LPDDR5X memory with error-correction code (ECC) with up to 1TB/s of memory bandwidth
- > 900GB/s NVLink-C2C

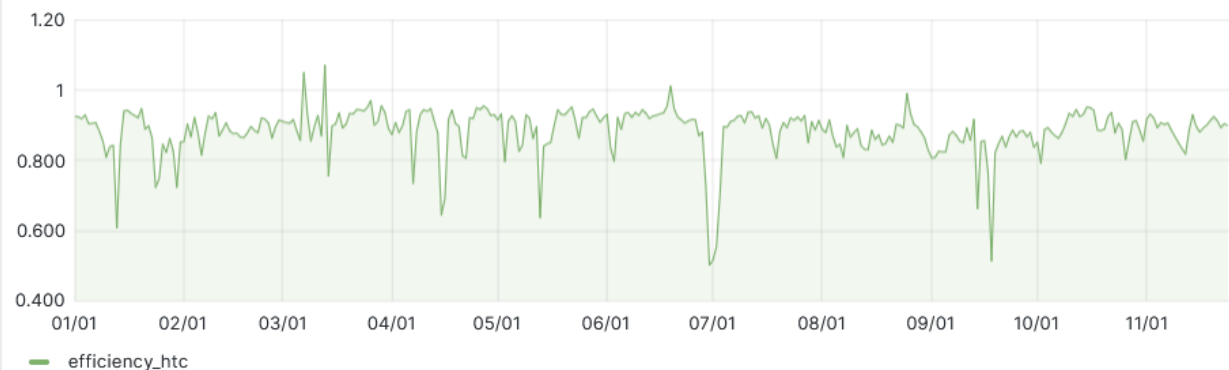
CPU ATLAS@T1 2023

atlas

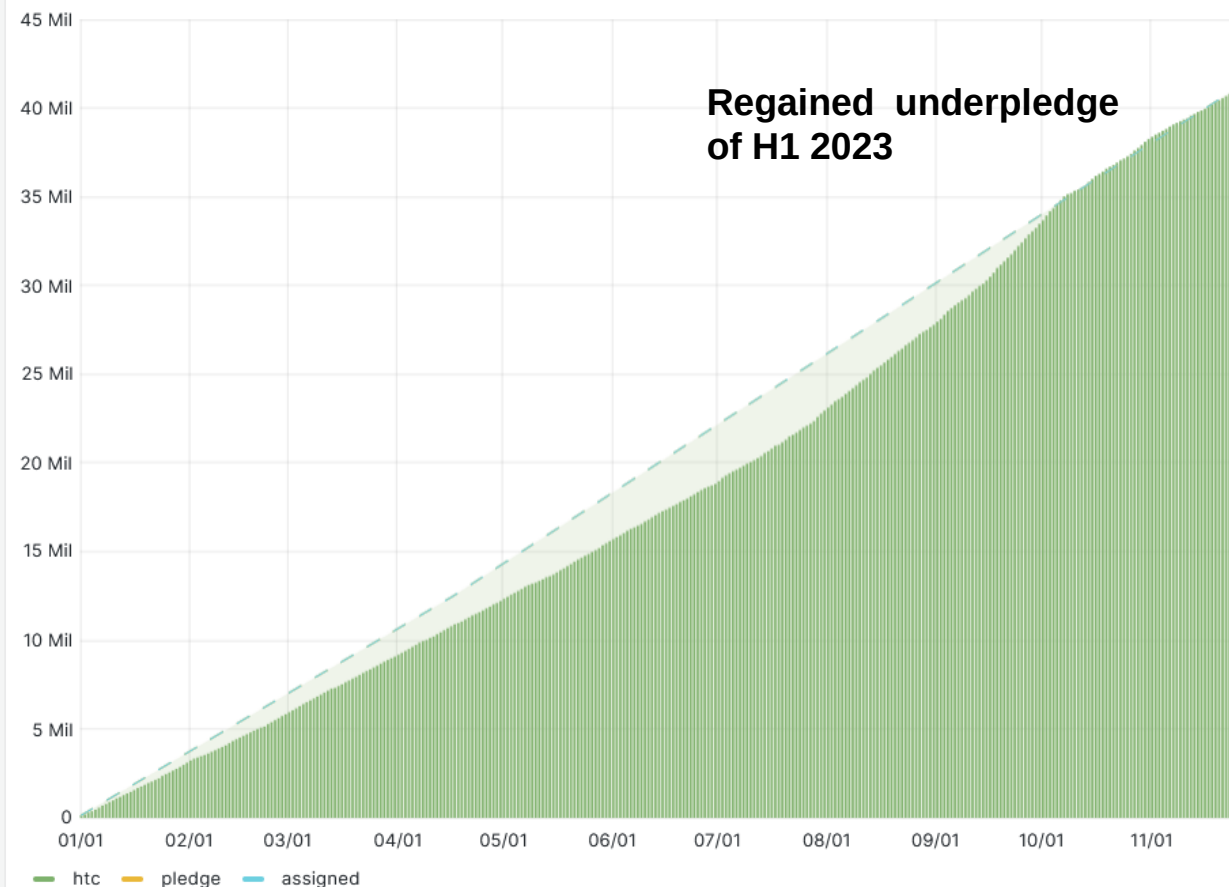
hs06_atlas



Efficiency_atlas



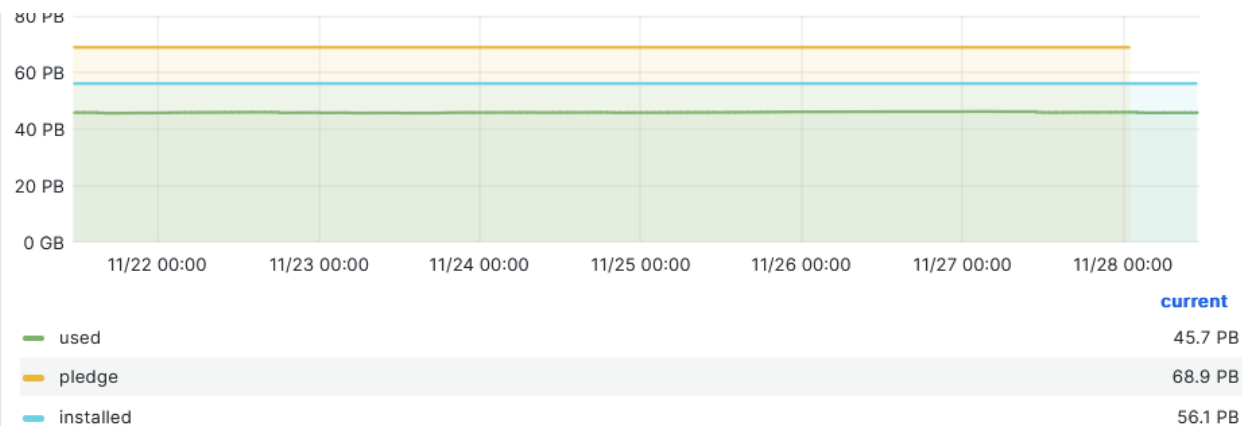
hs06_atlas cumulative



Storage Usage @T1 ALL VOs – no cloud

Space usage

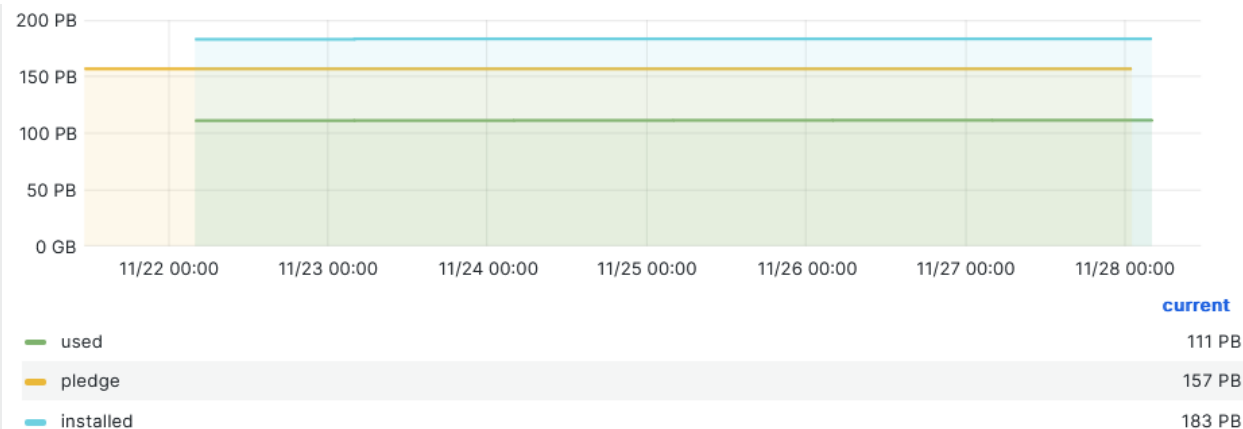
Disk space usage



DISK

- 12PBN underpledge
- Missing 2022 and 2023 resources
- Extending maintenance contracts for older systems that should be dismissed

Tape space usage

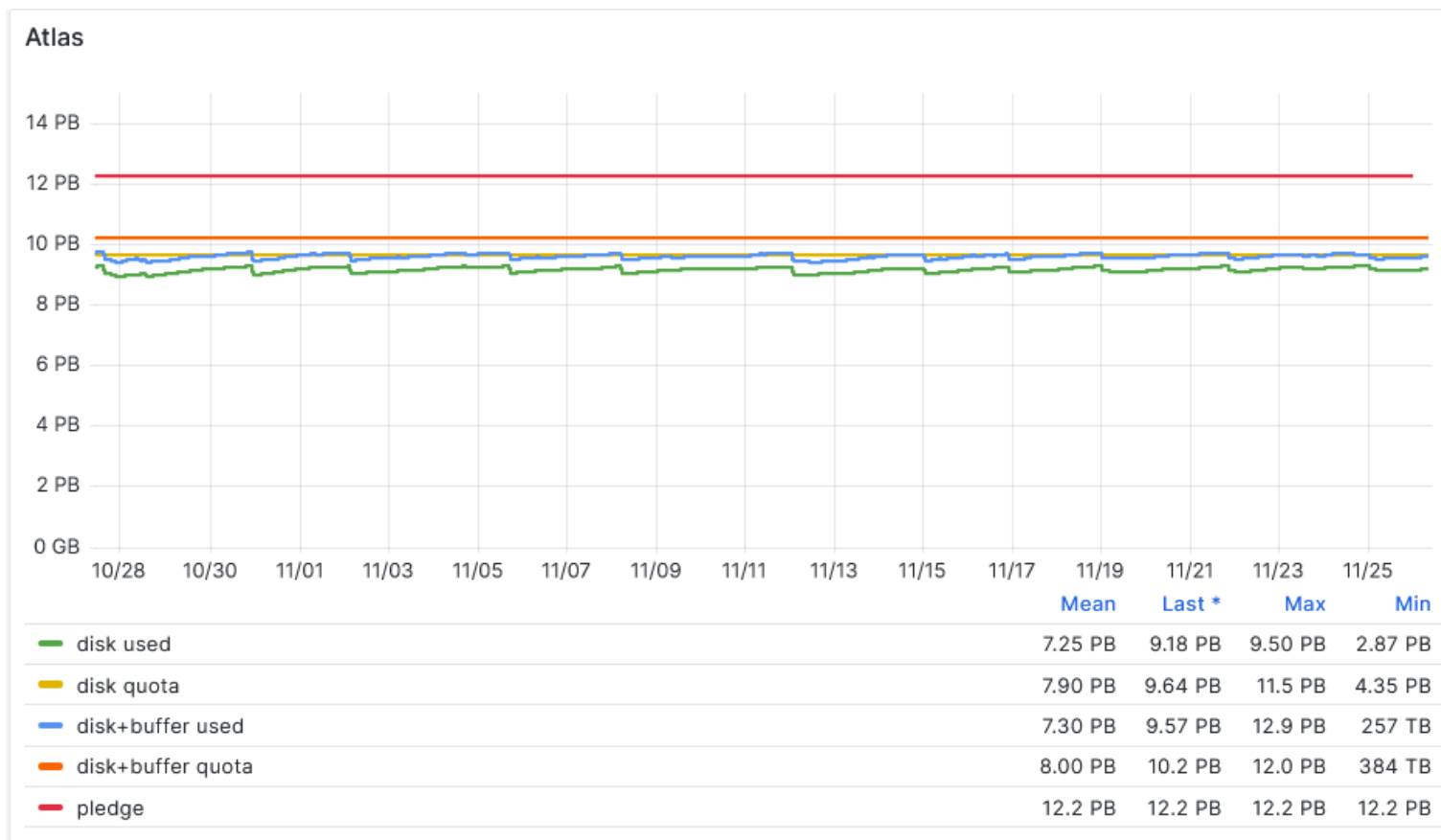


TAPE

- Installed greater than the pledge by 30PB
- due to Overpledge2023 + REPACK needs
- New tender under preparation 30PB (2023)

DISK ATLAS@T1 2023

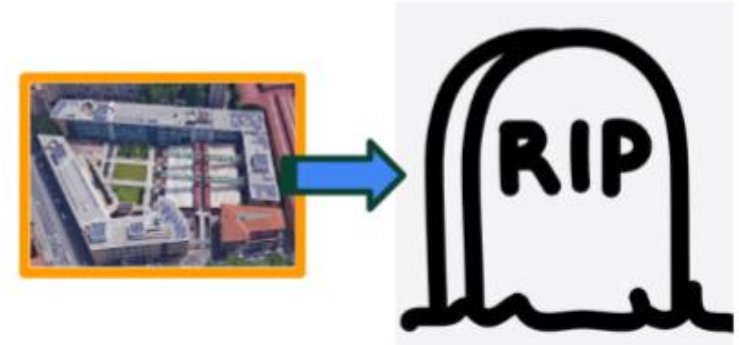
- 14PB della Gara 2022 mai arrivati
 - Il sistema LENOVO non ha superato il collaudo e non riusciamo ad assegnare la gara
- Abbiamo assegnato un AQ da 80 PB 2023-2024 finanziato con PNRR
 - Aggiudicato a Hitech Distribuzione Informatica
 - Contratto non ancora disponibile
 - Ordinato (64PBN) non ancora deliberato
 - Atti gara inviati in AC l'8/8/2023, poi OdF modificato il 31/08
 - Il vincitore è Huawei con sistemi OceanStore Micro 1500/1600
 - Installazione direttamente al Tecnopolo da inizio dicembre



ATLAS sotto pledge di 2PBN
 $\Delta_{\text{ATLAS}}(24-23) = 2.4\text{PBN}$

CNAF Tape Libraries and Drives

- **1 x Oracle SL8500**
 - **1 tape library with 16 tape drives T10000D** (8.5TB/cartridge)
 - 80PB installed, 64PB USED
 - Repack on the other libraries needed
 - After completion of repack this library will be dismissed
- **2 x IBM TS4500**
 - **1 tape library with 19 tape drives TS1160** (20TB/cartridge)
 - 102 PB Installed, 50PB USED
 - cannot be further extended due to physical constraints in the current room
 - This library will be moved to the new data center
 - **1 tape library with 18 tape drives TS1170** (50TB/cartridge) acquired and will be installed at new data center Q1 2024



38



Metropolitan Tape Area Network

- 2 libraries at CNAF
- 1 new library at the Tecnopole
- About **7 km** of fiber to connect the 2 datacenters
 - **yellow + red** paths
- 2 fiber pairs dedicated to extend the fiberchannel TAN
 - Brocade optics for 10km distance

Product Brief

Brocade® 32Gb/s LWL (10 km) SFP+

Optimized, Certified Optical Transceivers for Extending Service Provider and Data Center Networks

Overview

Today's enterprise data centers are undergoing an infrastructure transformation, requiring higher speeds, greater scalability, and higher levels of performance and reliability to better meet the demands of business. As speed and performance needs increase, optical transceivers—once considered a generic component of Fibre Channel switching technologies—have become an integral part of overall system design.

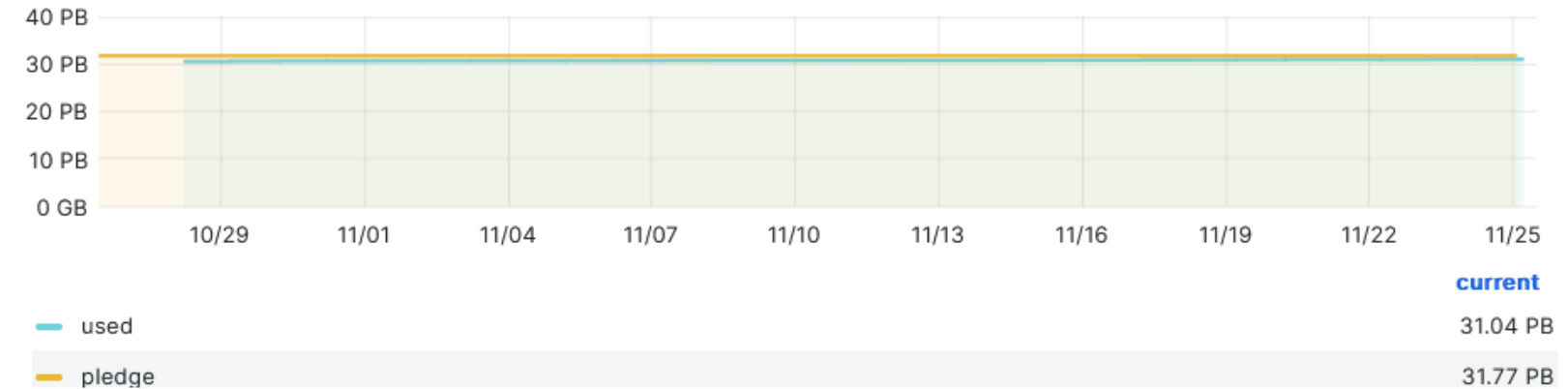
The Brocade® 32Gb/s Long Wavelength (LWL) 10 km SFP+, part of the

Highlights

- Provides high system reliability through rigorous qualification and certification processes.
- Leverages unique design parameters to provide the highest performance with industry-leading Brocade

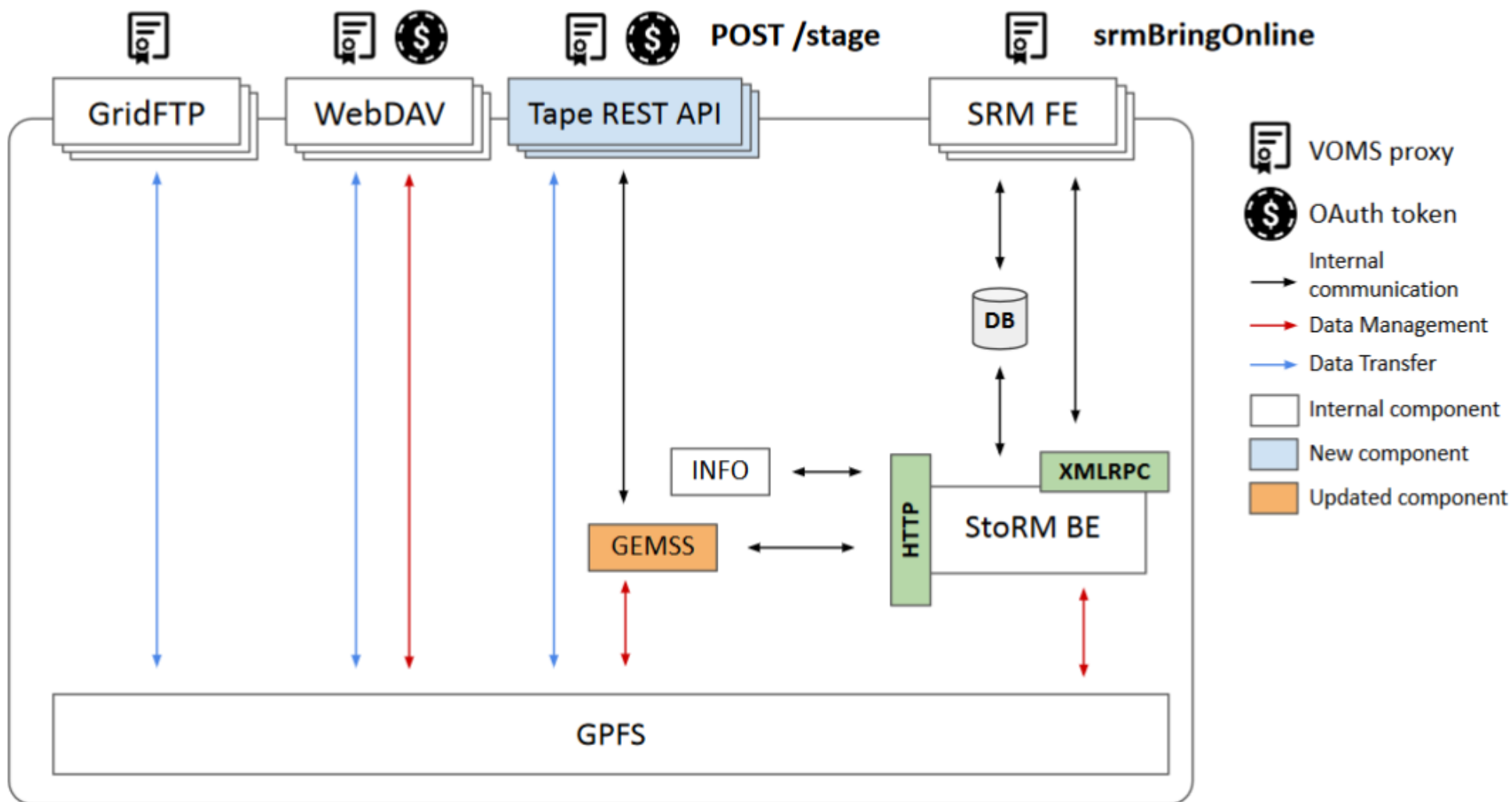
- Tender new Tape Library
 - Contract probably in January
 - Installed at Tecnopolo
- Just installed 50PB of tape cartridge in the IBM library
- Ongoing repack from the old Oracle to the new IBM that will be moved to the Tecnopolo
 - Repack will last until the end of 2024
 - About 130TB/day

atlas used vs pledge



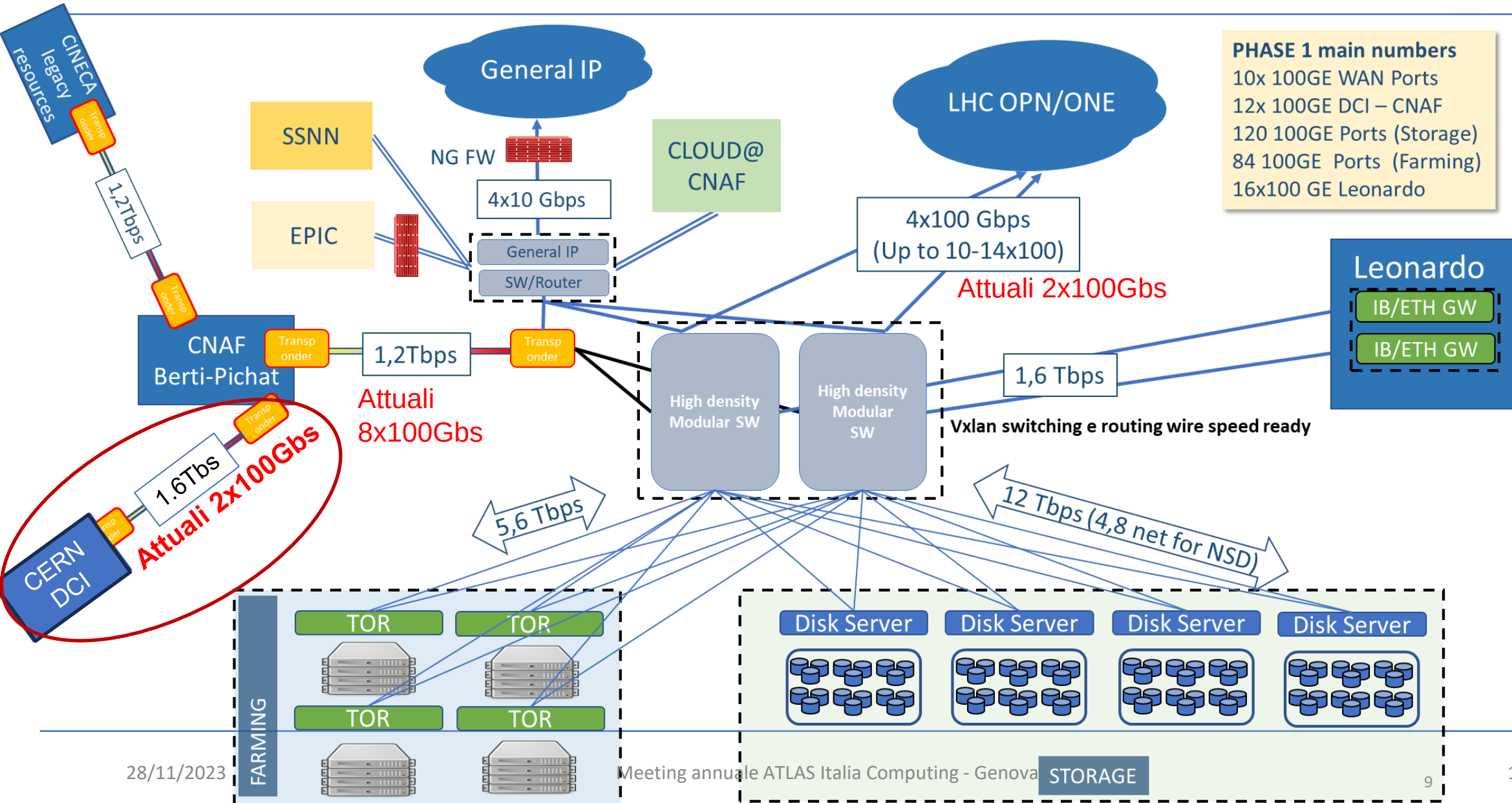
- No critical situation for tape at the moment
 - $\Delta_{\text{ATLAS}}(2024-2023) = 8.9\text{PB}$
- New tender - 30PB in January
 - Second round of 2023 acquisition
 - 20PB in old technology (20TB/tape), 10 PB in new technology (50TB/tape)

StoRM deployment: the Tape REST API scenario



- Current Storage SW in production
 - GPFS 5.1.2-11
 - StoRM BackEnd 1.11.22 (latest)
 - StoRM FrontEnd 1.8.15 (latest)
 - StoRM WebDAV 1.4.2 (latest)
 - StoRM globus gridftp 1.2.4
 - XrootD 5.5.4-1
 - ALICE CEPH updated to 5.5.5-1.el8
 - Ceph 16.2.6 (Pacific)
- **ATLAS StoRM Tape REST installed and configured**
 - NO tests yet (ATLASGROUPTAPE)
 - Implemented HTTP request/response for all the API endpoints defined in the WLCG specification
 - Packed the entire service in a Docker image (also as RPM)
 - Persistency with SQLite
 - Interaction with GEMSS for tape recal

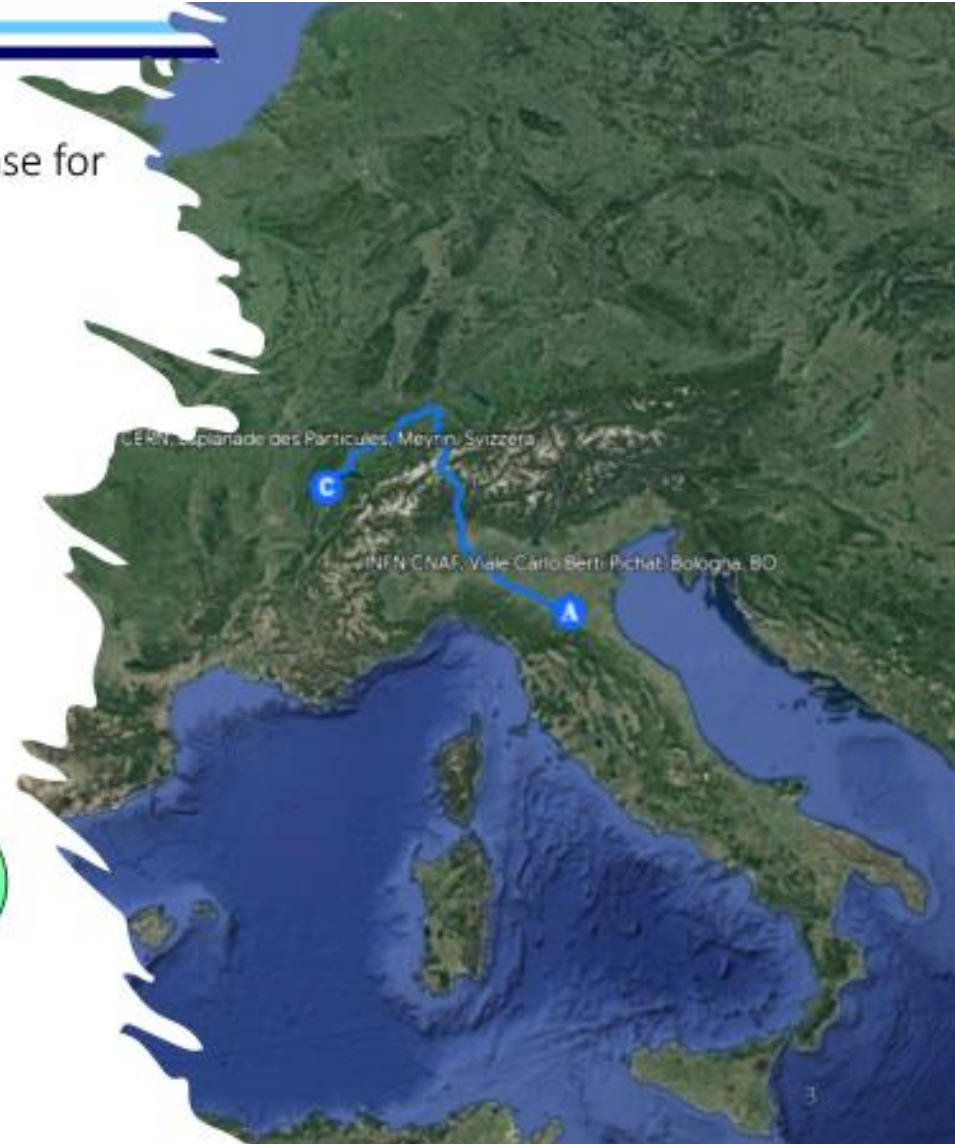
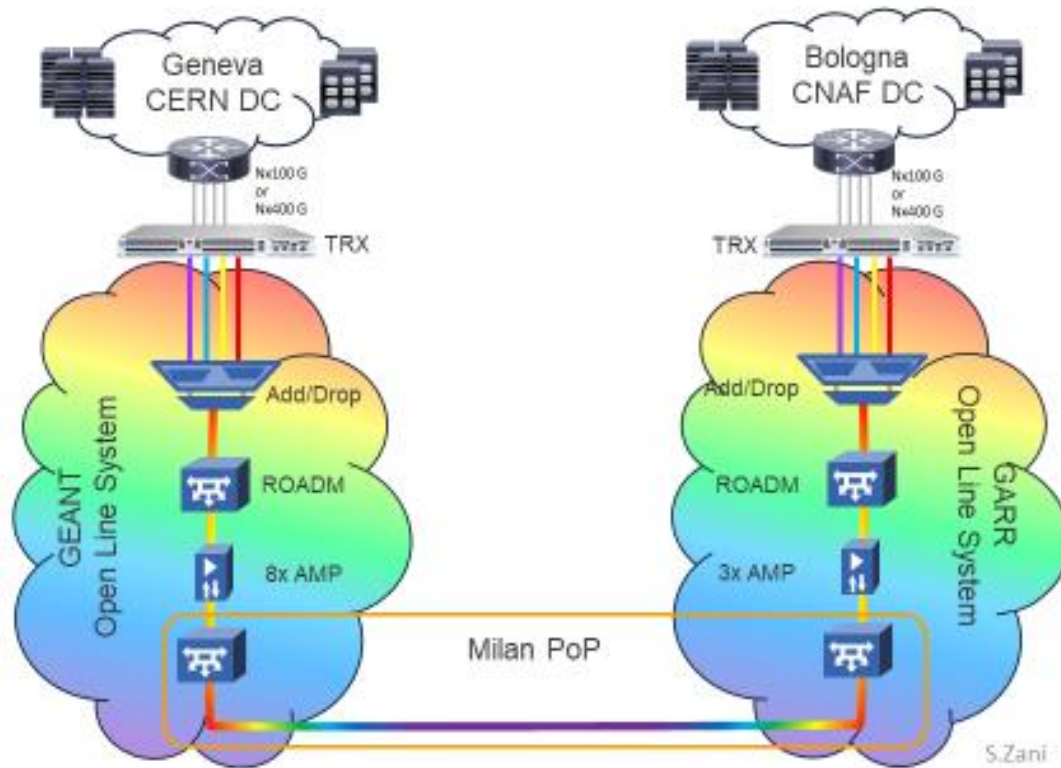
Networking Infrastructure



CNAF-CERN DCI

Schematic view of the DCI realized

Proposed in GEANT GN4-3 (WP7-T2) as a possible usecase for experimenting the multi domain Spectrum Connection Service at about 1000 km of distance.



Details in the
Stefano Zani talk at
LHCONE ws:

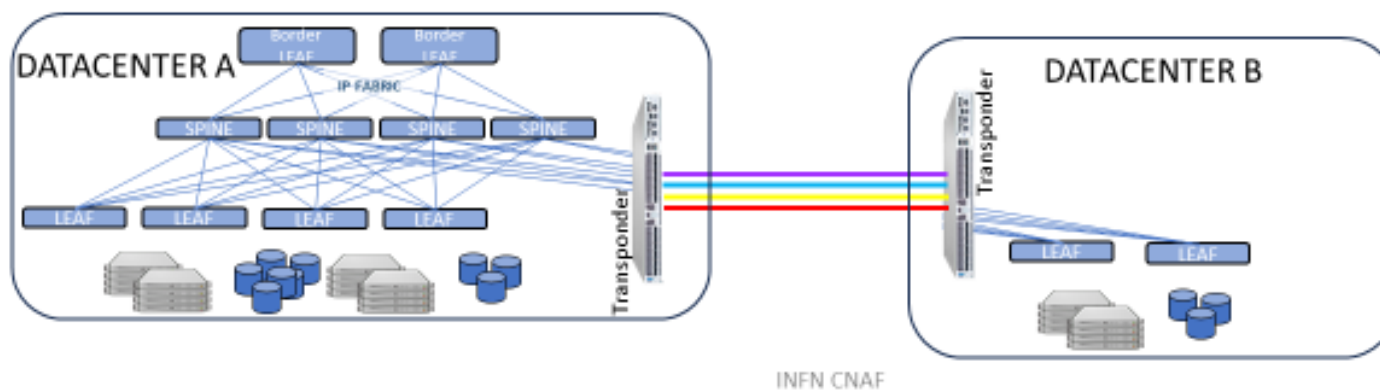
<https://indico.cern.ch/event/1280363/>

New opportunities

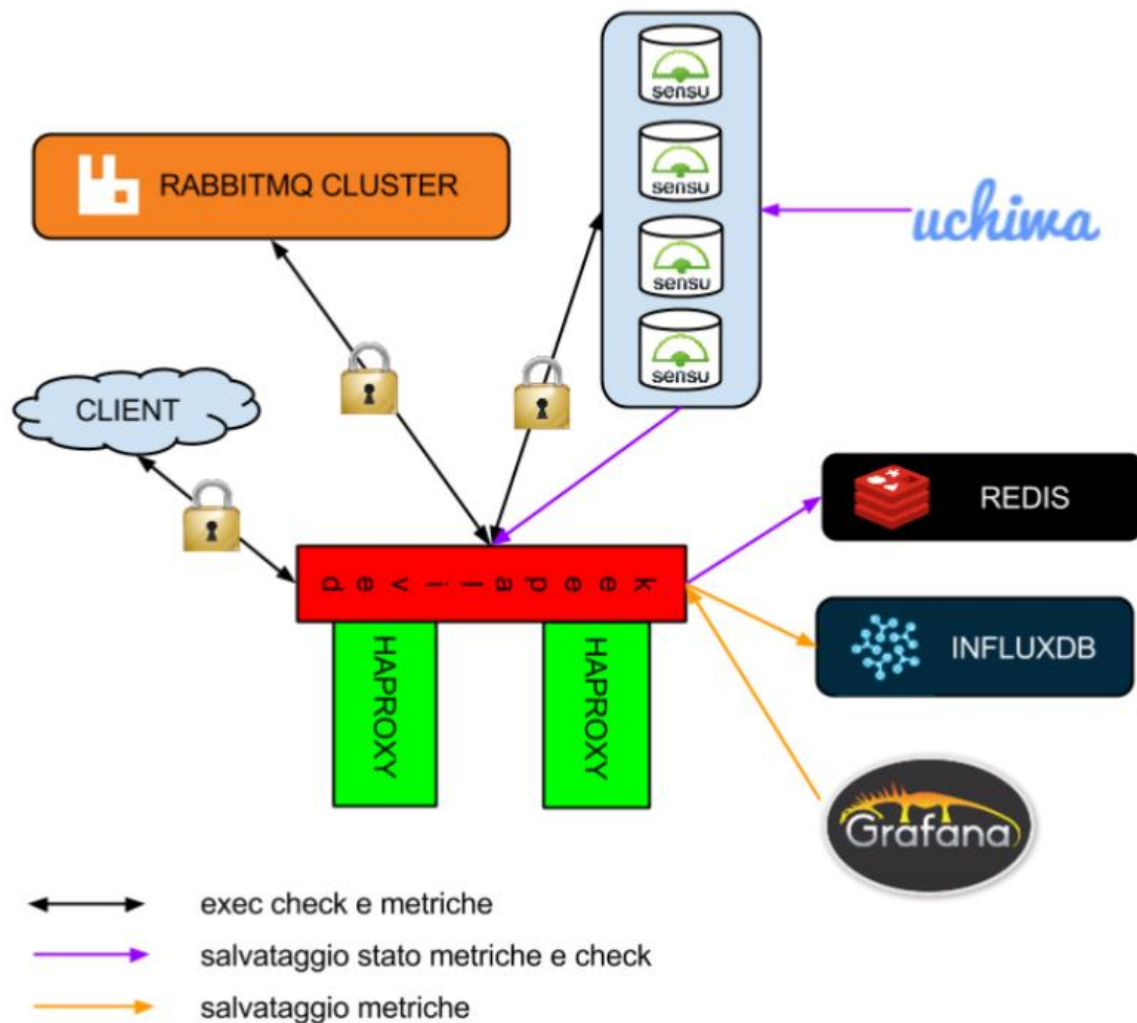
Considering that, this kind of DCI gives the possibility to connect up to 16 x 100G Ethernet or 4x400G Ethernet interfaces to whatever device both sides having only 9.5 ms RTT penalty the possible applications could be many.

Some examples:

- A «Classic» BGP peering with a dynamic number of ports contributing to the total capacity - **Easy growing in capacity (Shure interest for future production OPN link CNAF)**
- L2 Stretching of specific networks or IP Fabric stretching between the DCs having for example Leaves located in different Datacenters using the DCI links between spines and Leaves

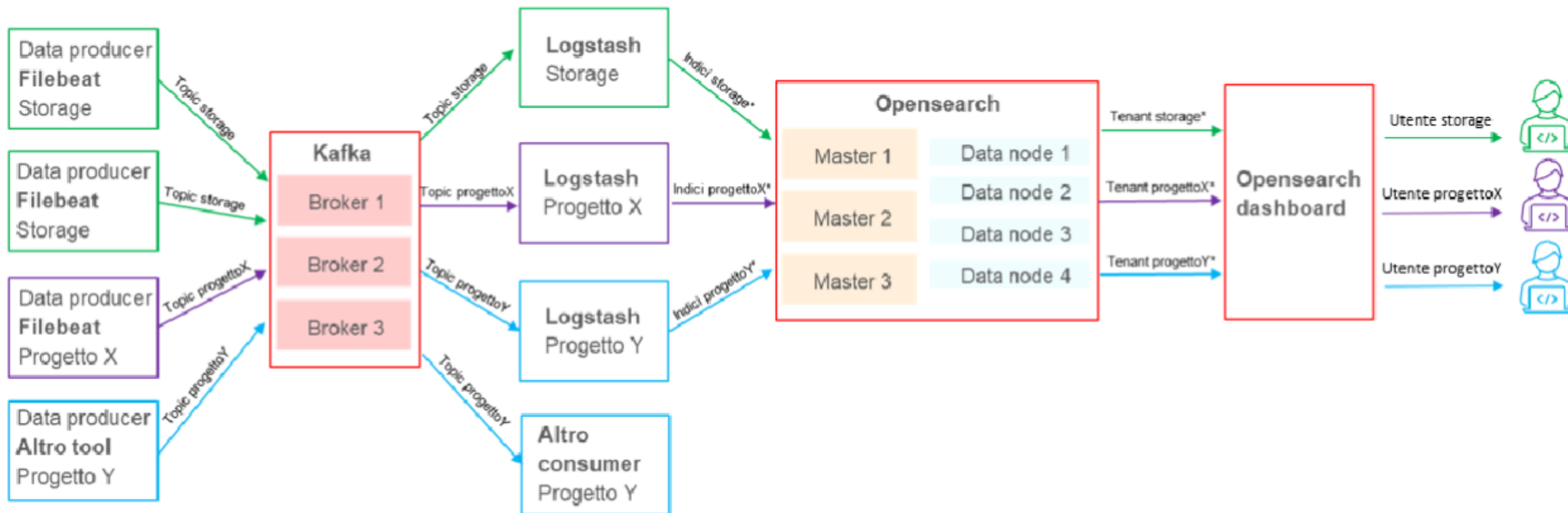


Monitoring Platform



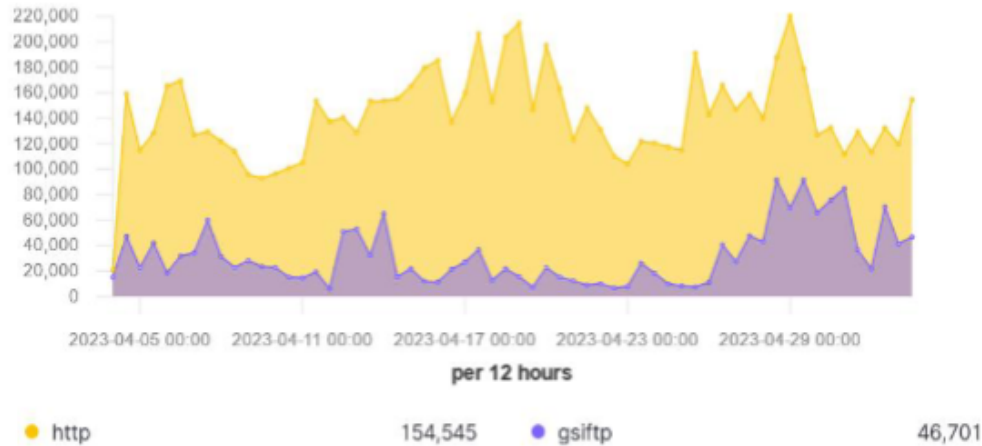
- Metrics, time series and alarms
- More details at:
 - <https://agenda.cnaf.infn.it/getFile.py/access?resId=0&materialId=slides&confId=733>

Big Data Platform@CNAF



- CCR WS 2023 talk e seminario al cnaf:
 - <https://agenda.infn.it/event/35255/attachments/103261/144896/bdpCNAF-29032023.pdf>
 - <https://agenda.infn.it/event/34683/contributions/197363/attachments/105682/148603/bdpCN AF-CCR-25052023.pdf>

Caso d'uso: analisi log dei servizi storage

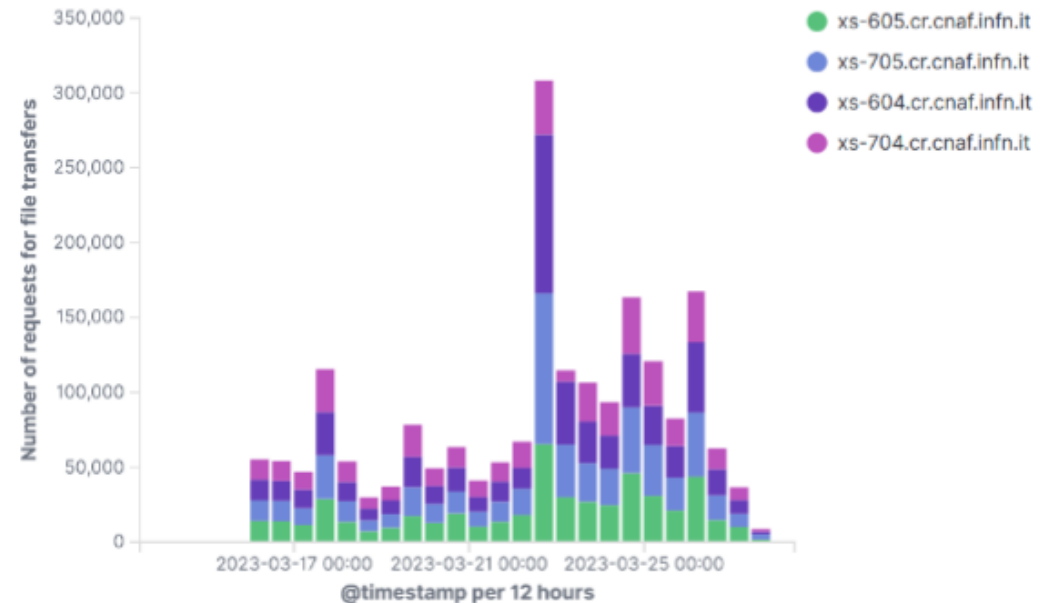


Numero di file trasferiti per protocollo (per monitorare il passaggio da GridFTP a StoRM WebDAV)



Numero e rate di trasferimenti http di ATLAS

Requests of file transfers to WebDAV



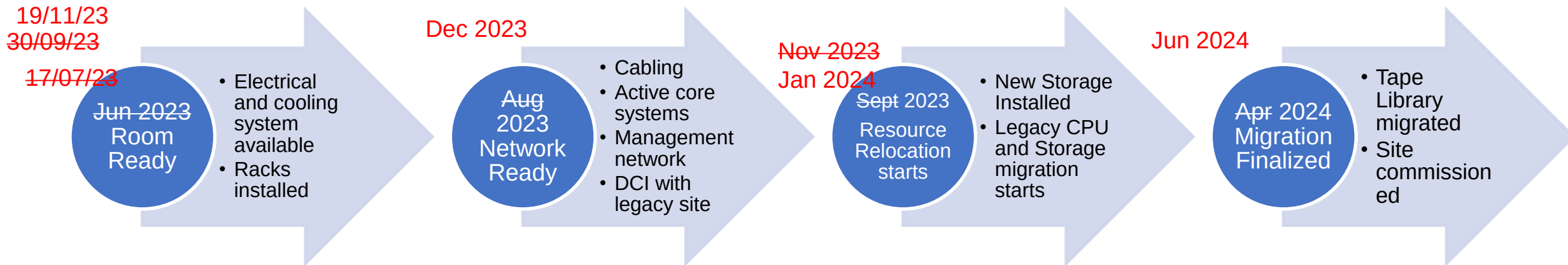
Distribuzione delle richieste di trasferimento file tra le 4 istanze di StoRM WebDAV dedicate ad ATLAS al CNAF





The new INFN Data Center at Bologna Tecnopolo

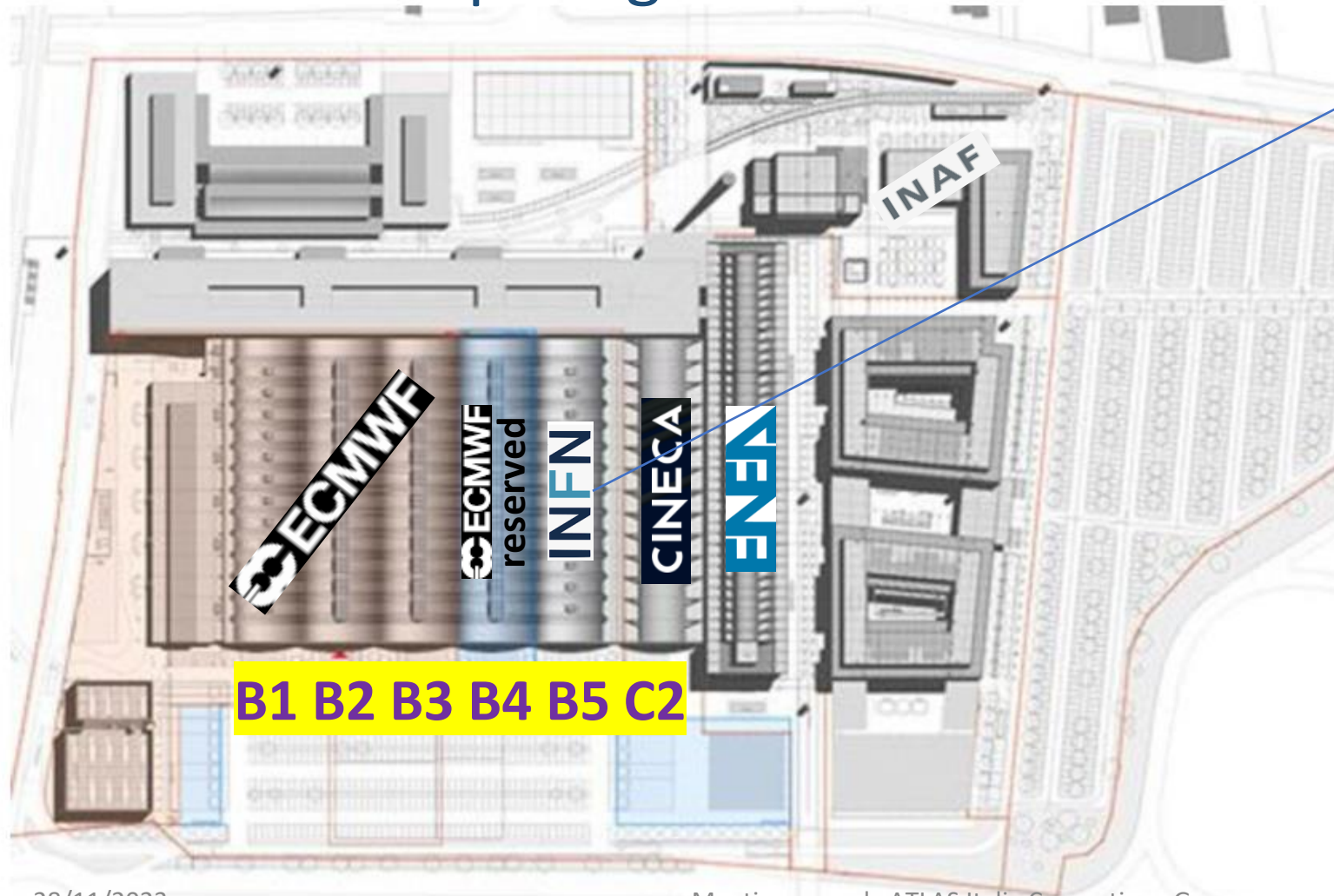
Live Relocation Timeline



- **Consegna sala rinviata innumerevoli volte per vari problemi tra cui amianto rinvenuto in un cantiere vicino (edificio ENEA)**
- **Il 04/12/2023 iniziano i lavori per il cablaggio di rete**
 - **6 settimane per completare il cablaggio**
- Apparati di rete attivi (core switch) consegna prevista per il 21/12/2023
 - Abbiamo già ricevuto apparti più piccoli come PoC
- Da metà gennaio 2024 iniziamo a trasportare il materiale dalla sede attuale
- Apparati di prova per lo storage in arrivo la prima settimana di dicembre
- Migrazione live
 - Down solo per tape library(ies)

What can the Tecnopolo host?

The computing infrastructures

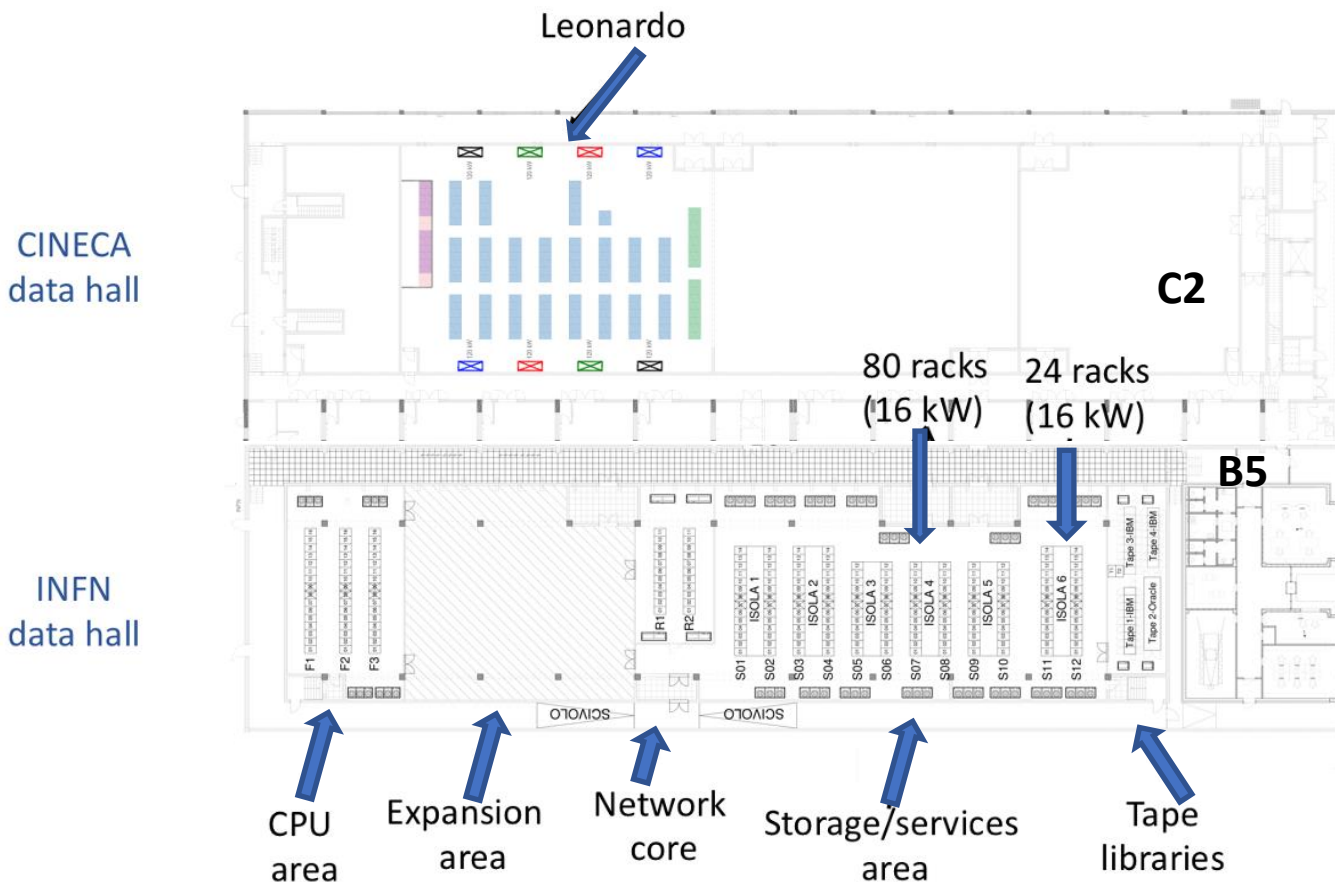


Each of the 6 “botti” (barrels) is
~5000m² of usable IT space



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

CNAF and CINECA data halls



DLC 80kW



- The new CNAF Datacenter will feature the following main areas
 - High Density – 2-3 rows for 80kW racks
 - Low density – 80+24 16kW racks
 - Expansion area
 - Tape libraries areas
 - Up to 4 libraries
- The CPU area can host up to 3MW of CPUs via 42 DLC high density racks
- The low-density area will be used to host
 - Storage systems
 - CNAF Cloud Infrastructures
 - ISO certified Cloud racks
- Cooling
 - Air cooled Cold Corridor aisles
 - Direct Liquid in High Density
- 3+1 redundancy in all the infrastructure facilities

The cooling system and the PUE

- 4 central refrigerator Units
 - 3+1 redundancy
- Chilled water 19-26 °C for the low density air cooled racks
 - 2 MW Chillers
 - Total/partial free cooling is possible
- Warm water 37-47 °C for DLC racks
 - 2,25 MW Chillers
- To be doubled in the second phase
- **High Density CPU Area**
 - 4 CRAH - 200 kW each (3+1)
- **Network Area**
 - 4 CRAH - 75 kW each (3+1)
- **STORAGE Area**
 - 16 CRAH - 200 kW each (12+4)
 - Cold corridor aisles
- **TAPE Area**
 - 4 CRAH - 25 kW each (3+1)

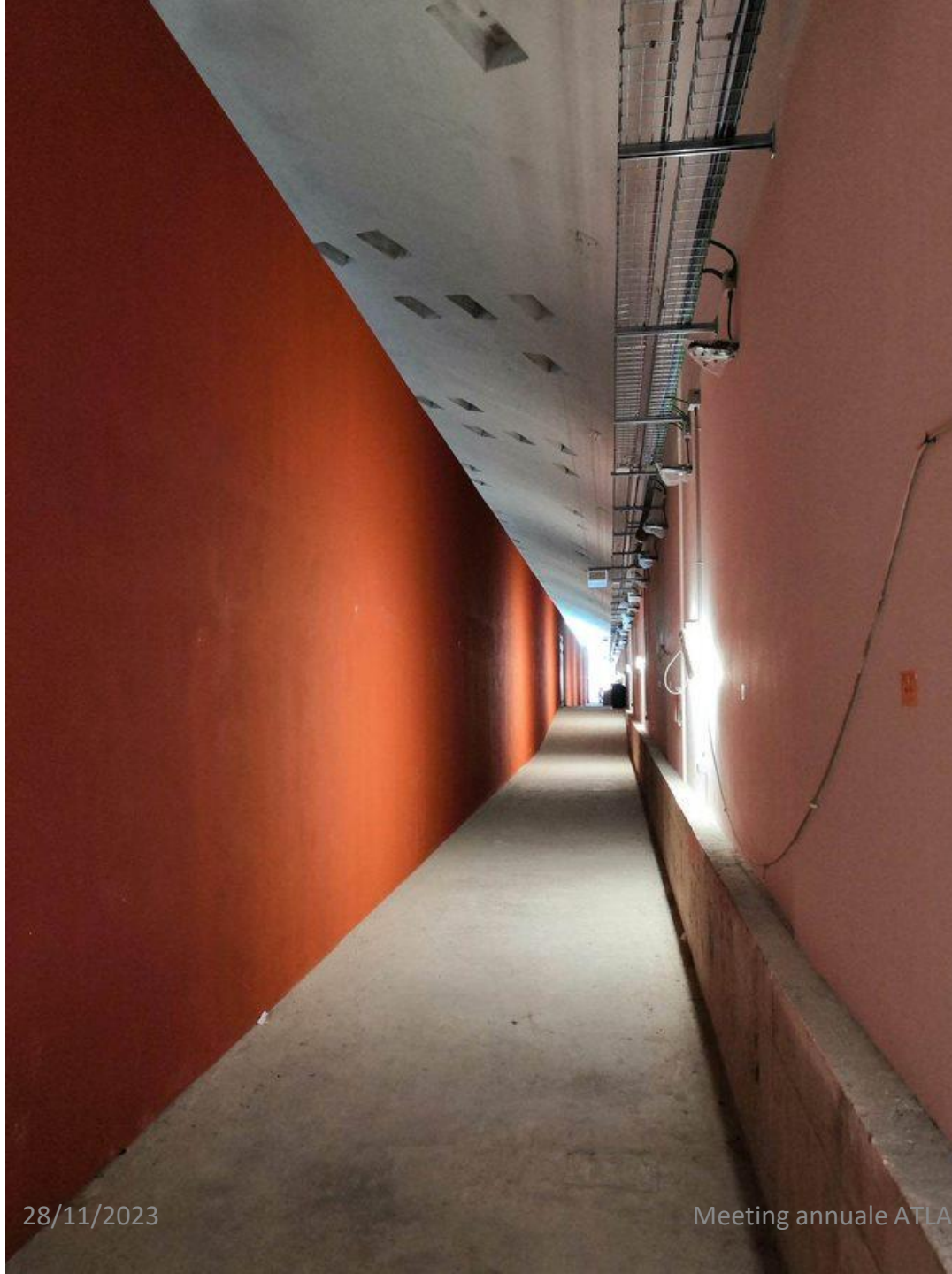
$$\text{PUE}_{\text{DLC}} \approx 1.08$$

$$\text{PUE}_{\text{Tot}} \approx 1.2 - 1.3$$

23 Novembre 2023

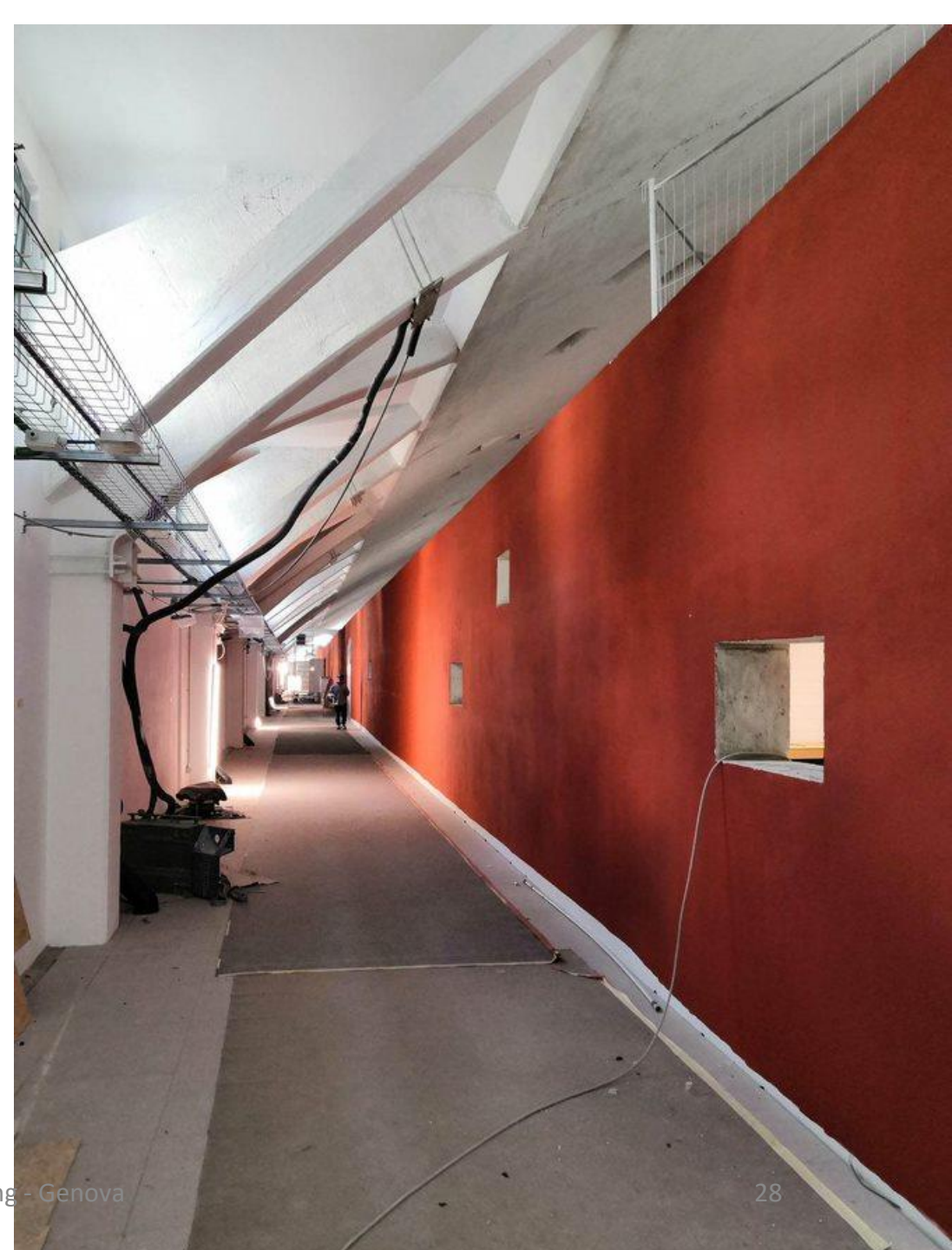






28/11/2023

Meeting annuale ATLAS Italia Computing - Genova



28



Storage 23/11



Rete 23/11



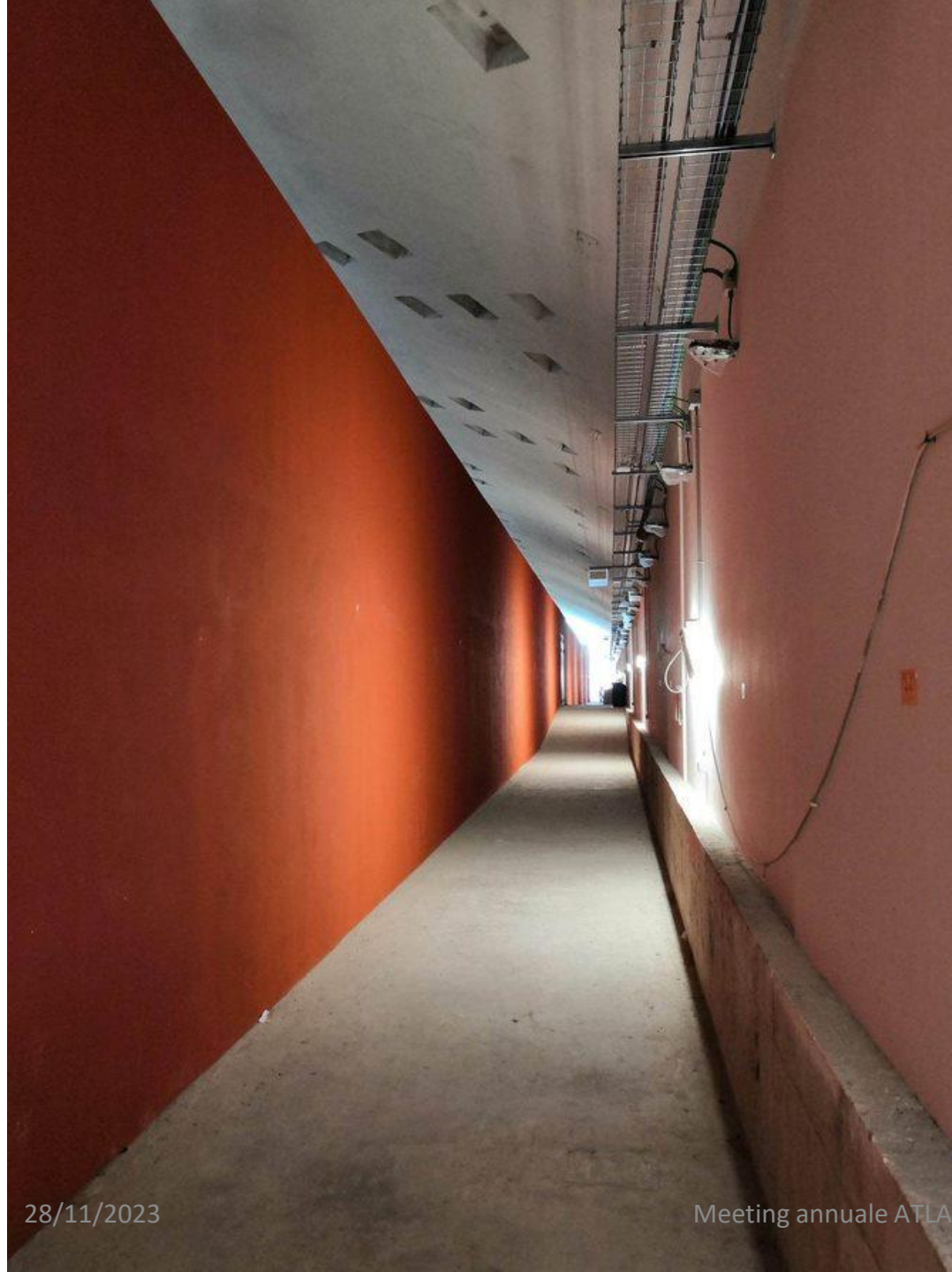
Alta densità 23/11



Tape 23/11

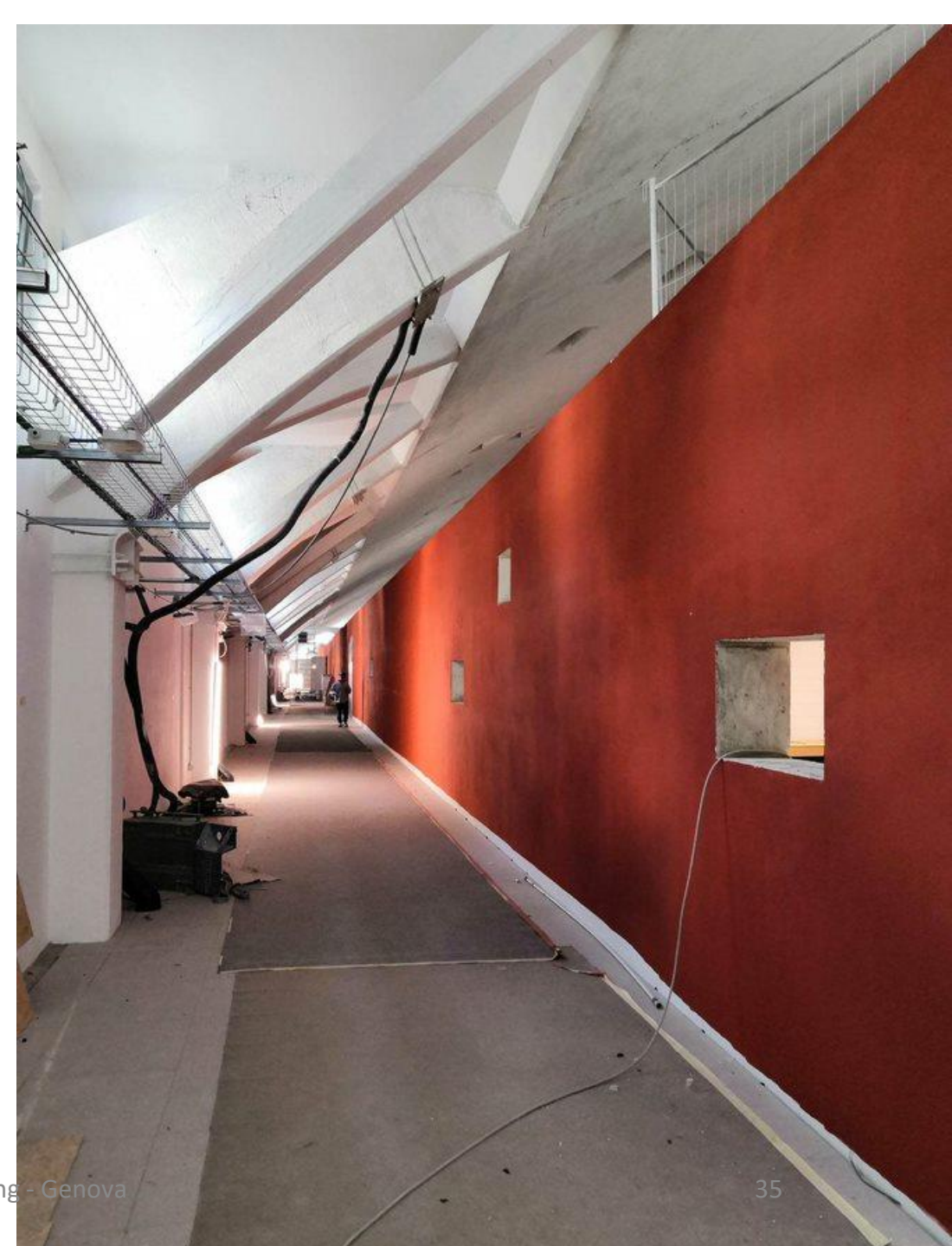
20 Aprile 2023





28/11/2023

Meeting annuale ATLAS Italia Computing - Genova



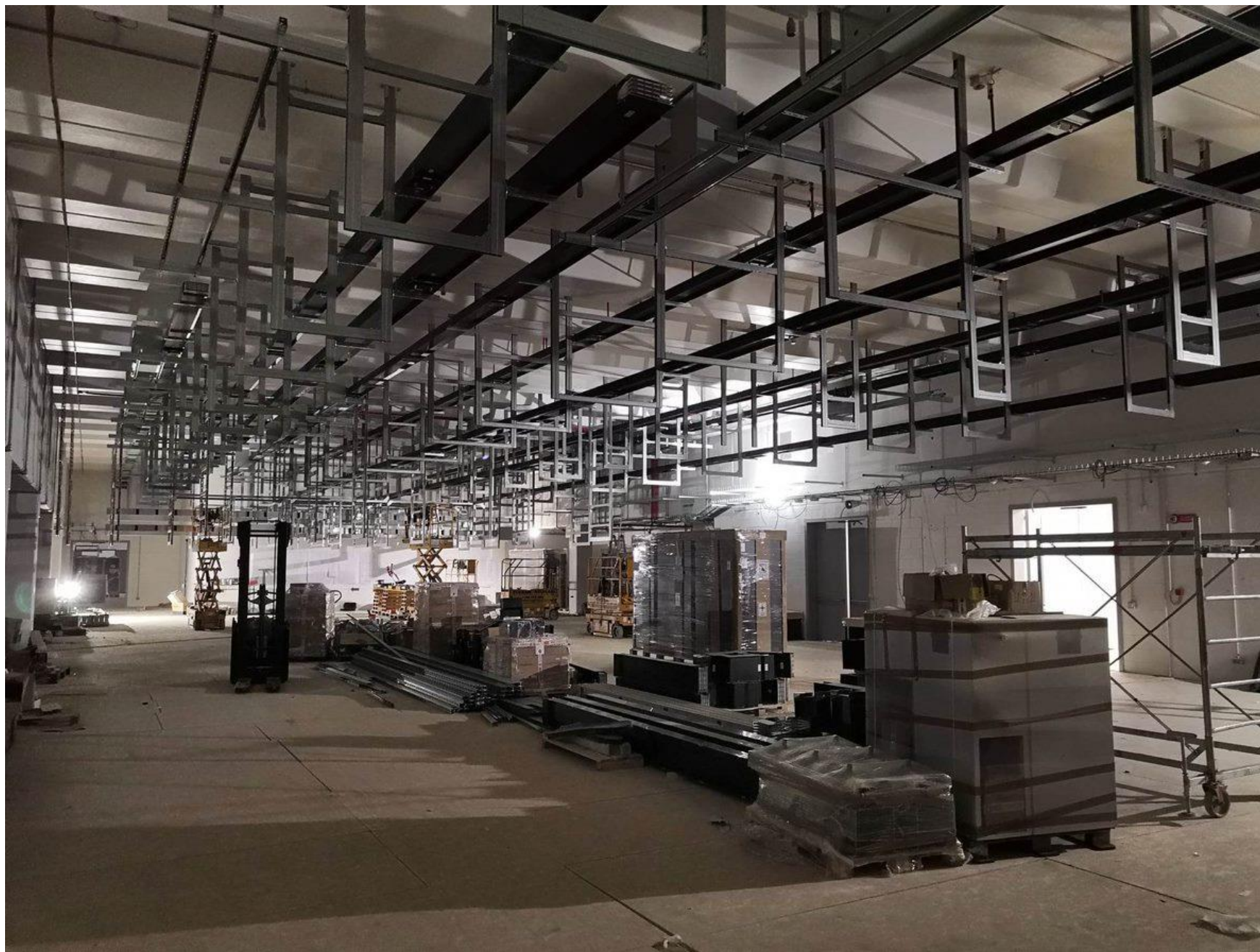
35

Zona Tape



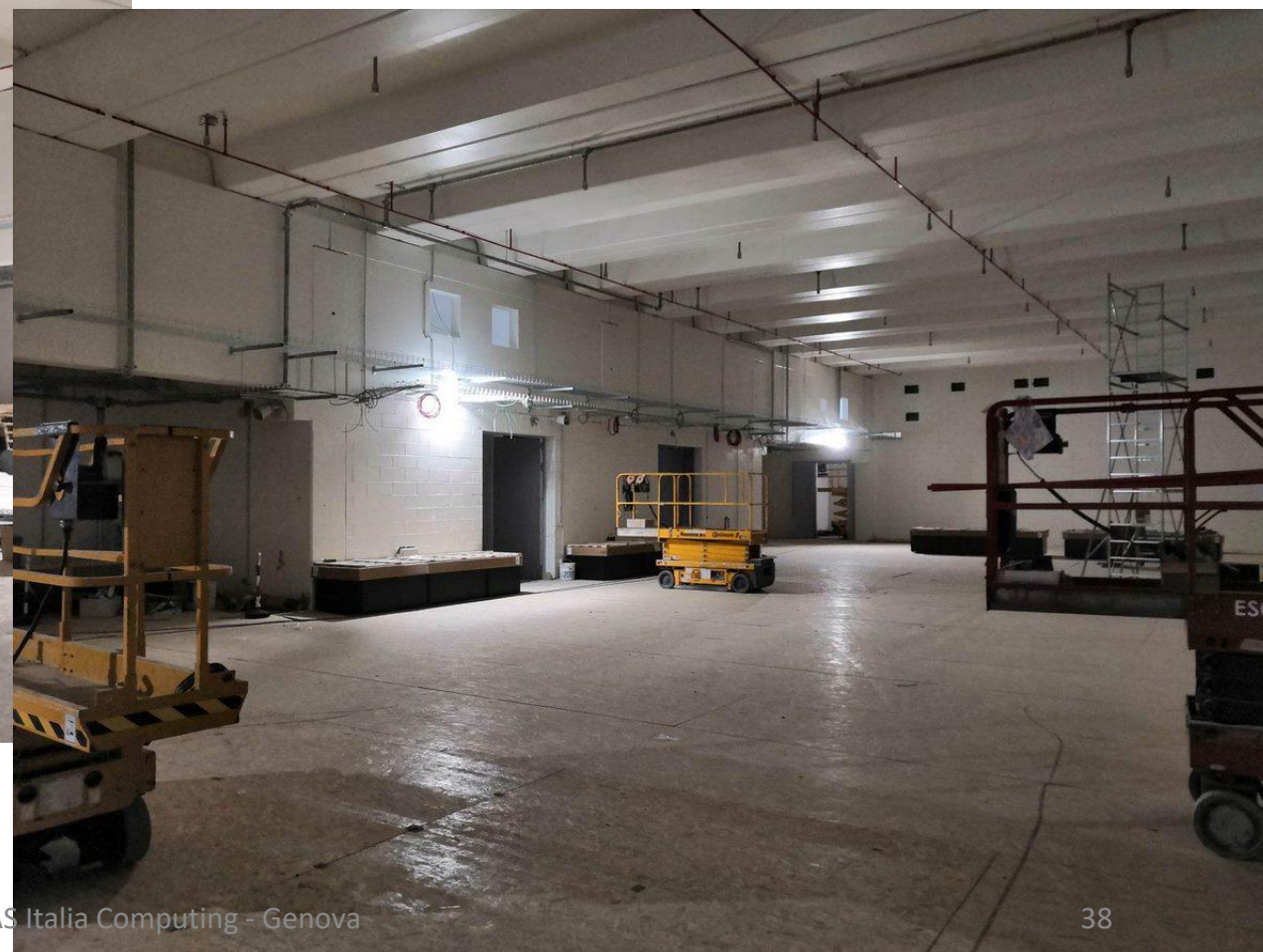
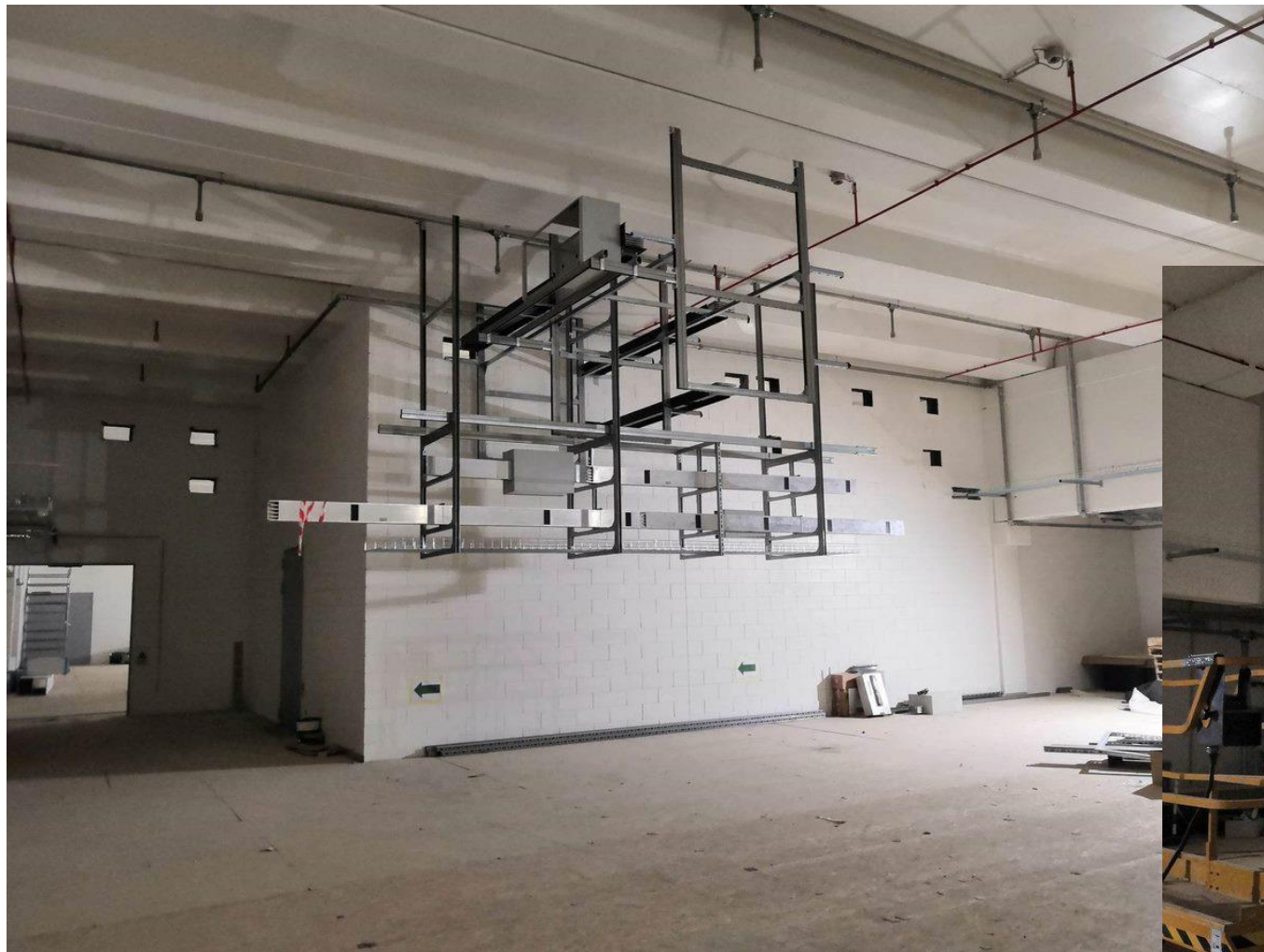
Zona Storage



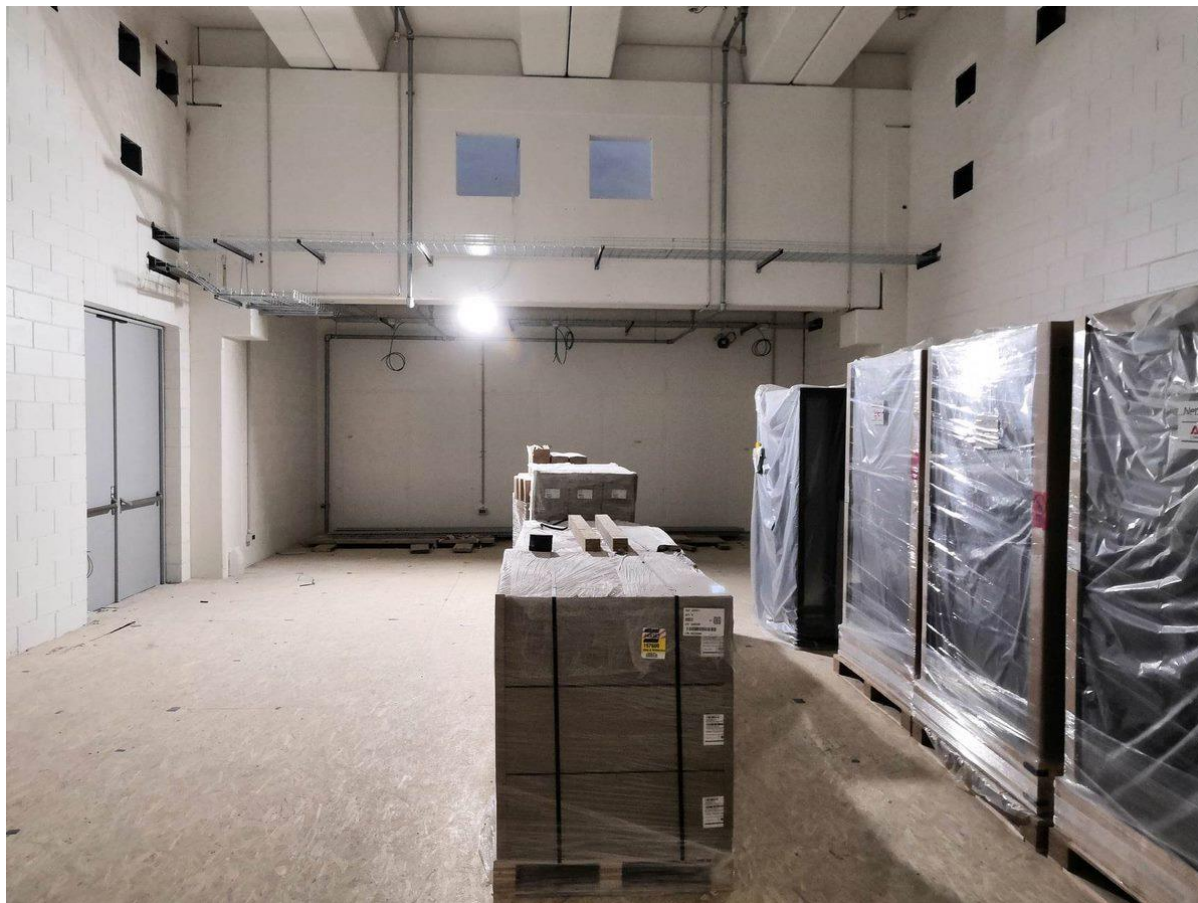


Storage 26/05

Zona Storage con Blindo



Zona Rete



Zona CPU



Zona CPU
Sotto il pavimento



Zona Espansione



28/11/2023

Meeting annuale ATLAS Italia Computing - Genova

40