

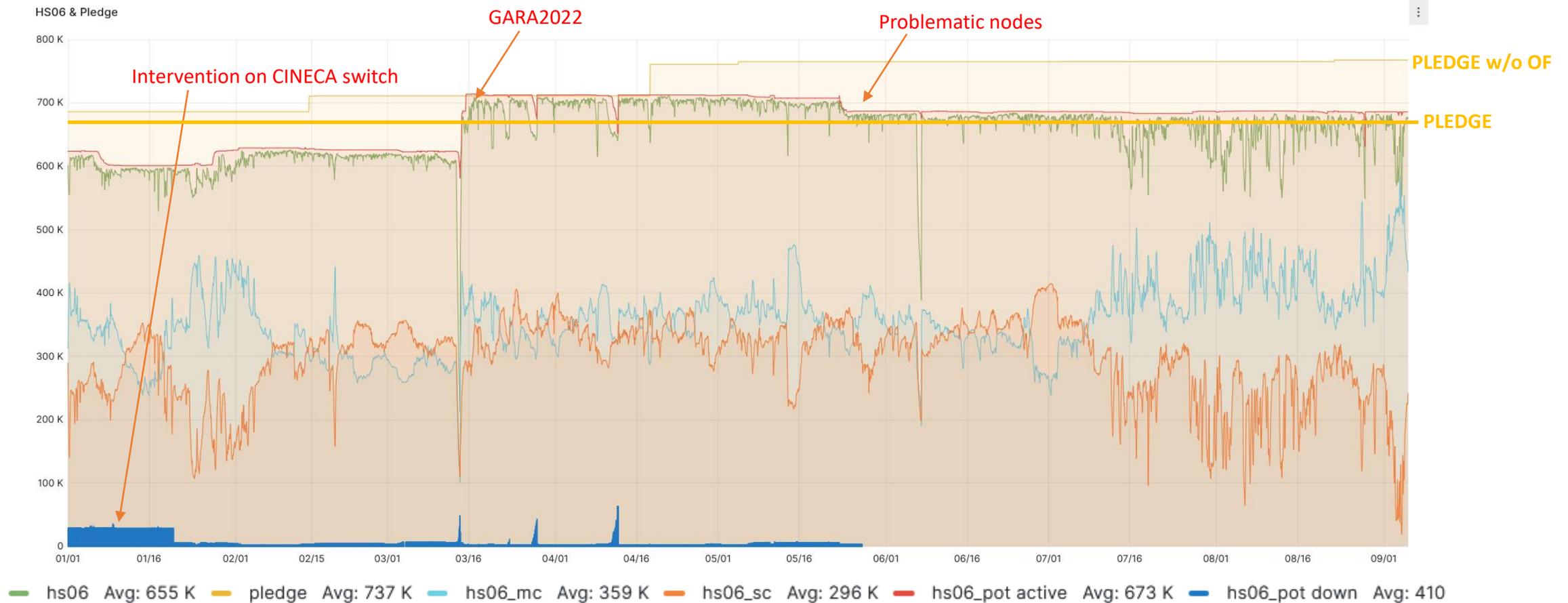
Referaggio risorse 2024

Stato e Risorse Tier1 e Cineca e Gare PNRR

D. Cesini – INFN-CNAF

CPU - Farm

- Pledge 2023: 660kHS06 (w/o OVERLAP- 770k)
- Gara CPU 2022 installata in Mar/Apr 2023

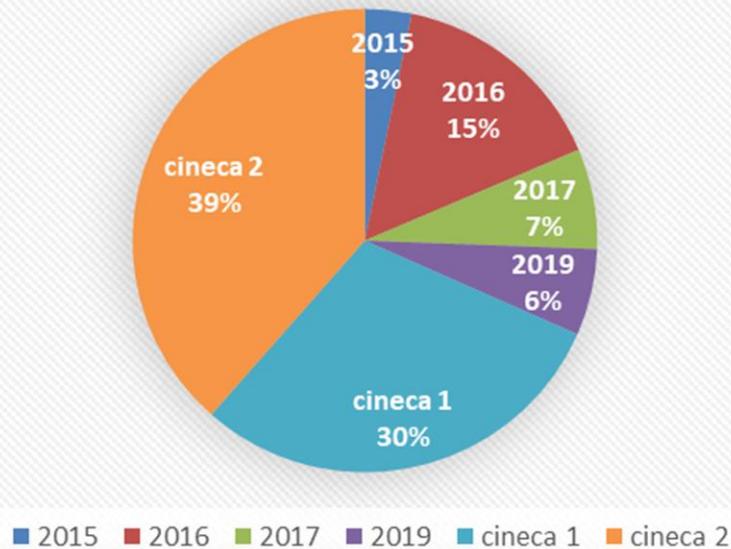


Composizione farm

- Siamo leggermente sopra pledge 2023: **686k HS06**
- 70% delle risorse T1 sono fornite dal CINECA

Leonardo GP node power estimate:
above 2k HS06, possibly >3k HS06

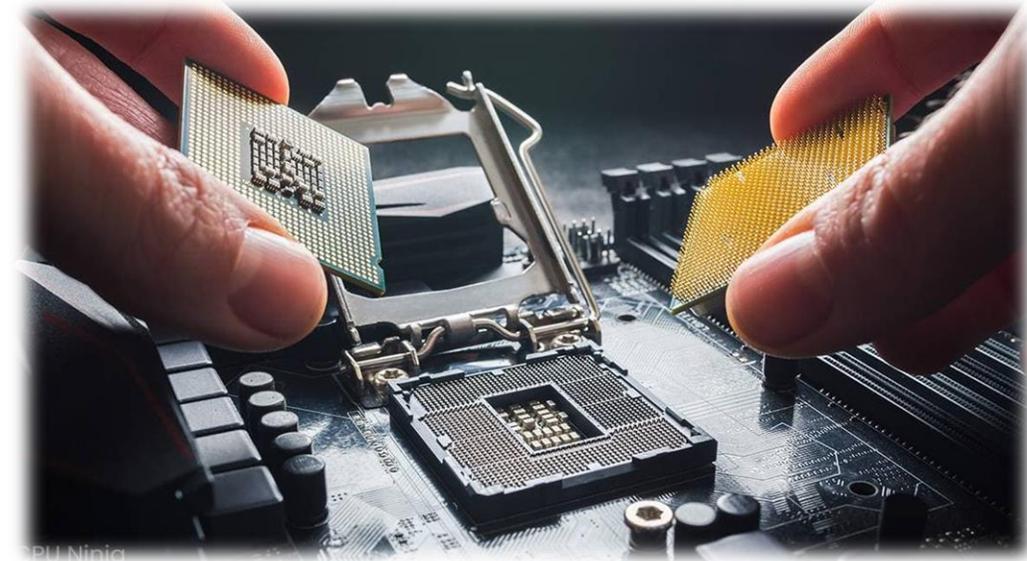
farm power per tender



Rimpiazzi e pledge 2023-2024
dovrebbero arrivare da Leonardo GP
con max 300 nodi nel 2024,2025, 2026

Rimpiazzi

Grazie a Leonardo
potremo dismettere
almeno 198kHS06



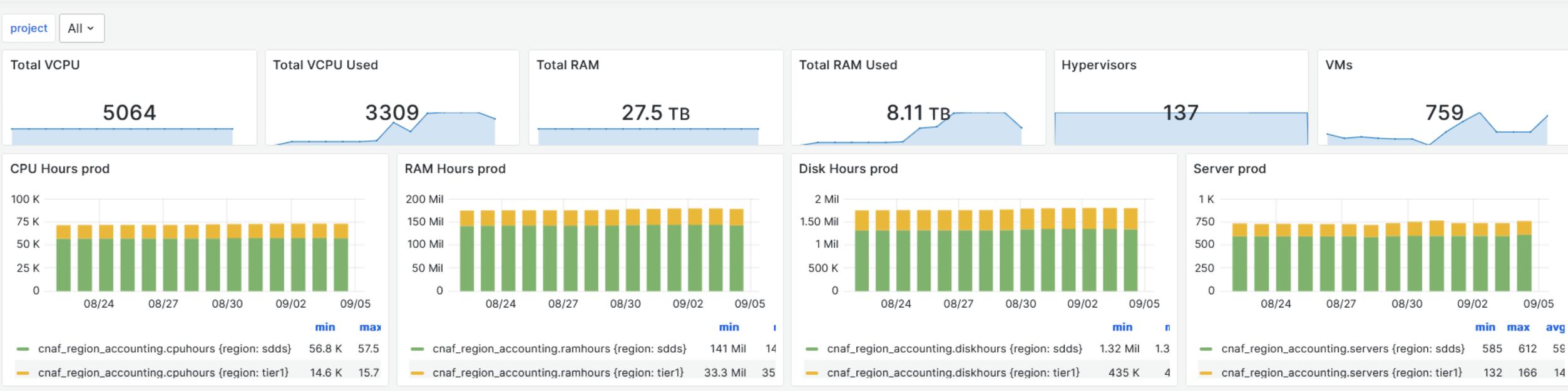
- CPU 2015: 54kHS06
- CPU 2016: 92kHS06
- **CPU 2017**: 42kHS06 possiamo tenere accese finché vanno, ma non portiamo al tp
- **CPU 2019**: 36kHS06 – possiamo portare al tp

Leonardo@CINECA

Set-up Leonardo GP per pledge 2023-2025

- Aggiornamento da riunione del 06/09
- Situazione molto, molto, molto complicata a tutti i livelli perchè.....
 - ...ve lo dico a voce
- Strategia che stiamo investigando:
 - WN creati tramite job SLURM “infiniti” che istanziano macchine virtuali WholeNode gestite da noi
 - Immagini VM create da noi e disponibili via shared fs
 - PCI passthrough per scheda infiniband
 - IP pubblico su interfaccia infiniband
 - Accesso inbound/outbound via NVIDIA Skyway collegati direttamente a nostri apparati
 - Pilot da creare nelle prossime tre settimane parallelamente alla connessione fisica degli Skyway (momentaneamente con solo 2x100Gbs per assenza ottiche)

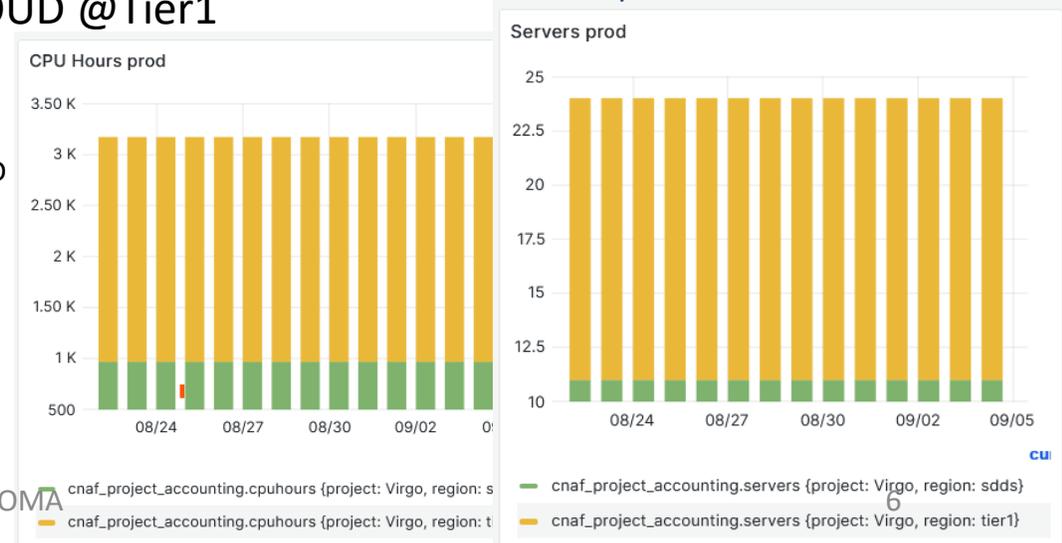
Stato Cloud@CNAF



Circa 100 tenant configurati, alcuni per esperimenti con pledge NON CLOUD @Tier1
 Tipicamente UI carrozzate su C@C per accesso interattivo

• VIRGO Low Latency Cluster on K8s

- AGATA-GAMMA
 - 2 istanze 8core 16GB RAM, 160GB HDD
 - 1 shared volume 2TB
- FOOT
 - 2 istanze 8core 16GB RAM, 160GB HDD
- QUANTEP
- MIBLAI/09/2023
- NTOF
 - 5 istanze 8core, 16GB RAM, 160GB HDD
 - 1 shared volume 1TB
- ASFIN
 - 1 istanza 32core, 64GB RAM, 280GB HDD



Stato esperimenti con Pledge Cloud

- Deployment su Cloud@CCNAF per servizi self-managed previa nomina di amministratore di sistema a utenti/ricercatori degli esperimenti
- Set-up iniziale fatto insieme al supporto
- Alcuni esperimenti accedono con interfaccia PaaS di INFNCLOUD

	CPU (HS06)	Disk (TB-N)	Costi (k€)
QUAX	100	130	17
AMS-02	200		2
HERD	1.000	100	22
SWG0	40		0
Fermi	1.100		11
AUGER	80		1
Cygn0	160	10	3
Totale	2.680	240	
Costi (k€)	27	29	56

➡ Fatto su C@C_t1 – via PaaS INFNCLOUD, assegnato metà disco

➡ Solo alcune VM (DB exp) C@C_t1

➡ FATTO via INFNCLOUD risorse al C@C

➡ Work in progress...

➡ Da contattare

➡ Da contattare

➡ Fatto – con Overpledge

- 2xCPU su backbone bari per 3Months
- 50TB in mirror su backbone (40GE+10CSN2))
- Via PaaS INFNCLOUD

ON-DEMAND SERVICES:

Virtual machine



Docker-compose



Run docker



SYS-ADMIN ONLY

Elasticsearch and Kibana



SYS-ADMIN ONLY

Spark + Jupyter cluster



SYS-ADMIN ONLY

HTCondor cluster



Working Station for CYGNO experiment



Project switched to admins/quax

CENTRALISED SERVICES:

<p>INFN Cloud object storage</p> 	<p>Notebooks as a Service (NaaS)</p> 	<p>INFN Cloud Registry</p> 
--	--	--

ON-DEMAND SERVICES:

<p>Virtual machine</p> 	<p>Docker-compose</p> 	<p>Run docker</p> 
<p>Run docker</p>  	<p>Elasticsearch and Kibana</p>  	<p>Spark + Jupyter cluster</p>  

Architetture alternative

- 2 nodi **ARM** Neoverse da 256 core ora disponibili
 - HEPSCORE preliminare: **3900**
 - Nodi di qualità e affidabilità ancora **non paragonabili** a x86
 - Assistenza solo on-factory, rischio di surriscaldamento con alcune configurazioni
 - Utilizzo possibile sia in batch che interattivo
- Prossimo passaggio: **RISC-V**
 - Schede molto meno performanti ma di sicuro interesse per la comunità'
 - Incaricato almeno un vendor "abituale" di indagare su possibili soluzioni

Disk storage in produzione

Installed: **53.64 PB**, Pledge 2023: **69.6 PB**, Used: **47.4 PB**

Storage system	Model	Net capacity, TB	Experiment	End of support
ddn-10, ddn-11	DDN SFA12k	10120	ALICE, AMS	12/2022 (20 spare hdd)
os6k8	Huawei OS6800v3	3400	GR2, Virgo	12/2023
md-1,md-2,md-3,md-4	Dell MD3860f	2308	DS, Virgo, Archive	11/2023
md-5, md-6 e md-7	Dell MD3820f	50	metadati, home, SW	11/2023 e 12/2024
os18k1, os18k2	Huawei OS18000v5	7800	LHCb	12/2023 Rinnovati fino a Gen. 2024
os18k3, os18k5, os18k5	Huawei OS18000v5	11700	CMS	2024 DEC 2023
ddn-12, ddn-13	DDN SFA 7990	5840	GR2,GR3	2025
ddn-14, ddn-15	DDN SFA 2000NV	24	metadati	2025
os5k8-1,os5k8-2	Huawei OS5800v5	8999	ATLAS	2027
Cluster CEPH	12xSupermicro SS6029	3400	ALICE, cloud, etc.	2027

Necessario estendere manutenzione
Da Rimpiazzare con AQ 23-24
Da spostare al Tecnopolo

Mancano 14PB da Pledge Storage 2022 – bloccati nella relativa Gara – Collaudo non superato!

Stato AQ Storage 2023-2024

- Fino a 80PB_N
 - Max 64PB nel 2023, max 16PB nel 2024
 - 63,87 E/TB-N IVA esclusa (77.92 E/TB-N IVA inclusa) – 35.07% di ribasso
 - Da installare al tecnopolo
 - Primo AS congiunto all'approvazione degli atti della gara
 - GE 12/09/2023
 - 2.4ME su Terabit, 1.6ME su ICSC, 1.6ME su FOE
 - 64PB-N che ci permetterebbe di:
 - Raggiungere il pledge 2023 (mancano 16PB) 43.6PB finanziati nel 2023
 - Dismettere i sistemi da rimpiazzare (35PB)
 - Sopperire alla mancanza dei 14PB del 2022
 - Avere abbastanza cache per spostare i dati dei sistemi al CNAF, partendo da quelli più vecchi
 - Avere capienza per eventuali richieste di ICSC

Stato Tape

- Pledge 2023: 158.3PB
- Liberi 20 PB (su cassette vuote, complessivamente sulle 2 librerie)
- Usati 103 PB
- Installato attuale: 130.5 PB
- Spazio quasi esaurito sulla IBM -> maggior parte delle scritture sulla Oracle in attesa dei nuovi nastri

Library	Tape drives	Max data rate/drive, MB/s	Max slots	Max tape capacity, TB	Installed cartridges	Used space, PB	Free space, PB
SL8500 (Oracle)	16*T10KD	250	10000	8.4	~10000	57.3	19.6
TS4500 (IBM)	19*TS1160	400	6198	20	2450	45.7	0.3

- Nuova Libreria sarà installata direttamente al Tecnopolo
- Ci stiamo attrezzando per avere una TAN metropolitana in modo da eseguire più agevolmente il repack della Oracle

Gare Tape

- **Gara Tape Library**

- Finita la fase di Gara, da approvare gli atti (Ottobre 2023)
- IBM SL8500 con drive nuova generazione (JF) da 50TB/tape + 2 switch TAN Brocade
- Aggiudicata a 477.3kE IVA esclusa su fondi ICSC (con 18 drive)

- **Gare nastri**

- 2023: 53 PB: pledge + overpledge 2023 (vecchia generazione JE 20TB/tape)
 - Gara completata, atti approvati – manca il contratto – in fase di stipula
 - Aggiudicata a POSDATA 10,679% ribasso → 388033EU IVA esclusa → 8.9EU/TB IVA INCLUSA
- 2024: Ulteriori 50 PB di nastro per repack dei dati dalla libreria Oracle
 - Gara da iniziare asap

Pledge Delta(24-23) da Preventivi

		Euro
delta cpu (HS06)	55624.16667	556241.6667
delta disco (TBN)	11773	1412760
delta tape (TB)	31299	312990

2281991.667

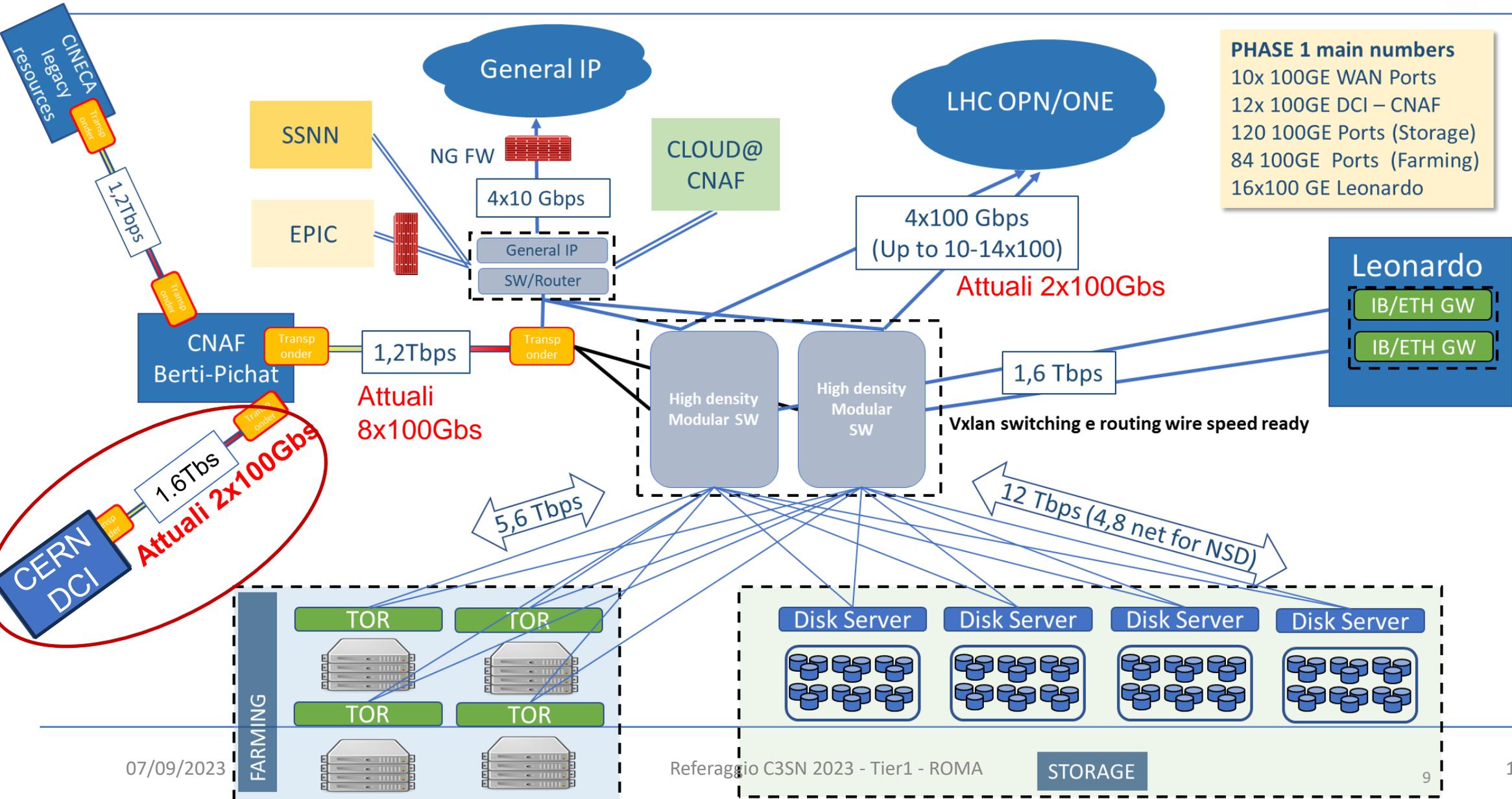
HPC

- Non abbiamo risorse “pledged” HPC
 - Piccolo cluster InfiniBand general purpose ormai super-obsoleto
 - 4 x NVIDIA V100 GPU
 - Non verrà portato al tecnopolo
 - Altre GPU recenti su cloud via progetti e ML_INFN
 - Secondo cluster HPC più recente ma in uso esclusivo al CERN
 - Contratto fino 2025
 - Nuove risorse via HPC bubbles con Terabit
- Uso partizione booster Leonardo opportunistica

Gare Rete

- Cablaggio strutturato di rete
 - Procedura completata
 - Circa 50% di ribasso, ditta N&C
 - In attesa di poter entrare in cantiere – poi 6 settimane per completamento
 - Richiesto alla GE il “sesto-quinto” per poter cablare du nuove sale tra B5 e C2 non inizialmente presenti nel Progetto esecutivo
 - Le fibre LEPIDA, GARR, Telecom etc atterrano qui
- Apparati core e infrastruttura management
 - Procedura completata da tempo
 - Sistemi ARISTA
 - Tempi per la consegna molto lunghi (Dicembre)
 - Accordo per avere in anticipo (Ottobre/Novembre) sistemi più piccolo ma con interface analoghe in modo da iniziare il setup

Networking Infrastructure



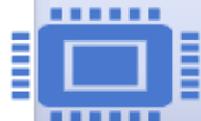
Live Relocation Timeline



- **Lavori edili al momento bloccati per amianto in cantiere vicino (edificio ENEA)**
- Dovrebbero sbloccarsi a giorni
- 6 settimane per completare il cablaggio
- Installazione apparati attivi di rete
- Migrazione live
- Down solo per le tape libraries
 - dismantle and re-assembling



HPC Bubbles



Nodo CPU

Min 112 core fisici (max 192)
RAM > 8GB/core DDR5
IB NDR 400G
20TBL + dischi di sistema



Nodo GPU

Come CPU + 4x NVIDIA H100 SXM5 con minimo 80GB e memoria HBM2e



Nodo FPGA

Min 32core
RAM > 512GB DDR4 o DDR5
IB NDR 440G
4 x XILINX U55C o 4 x TerasicP0701



Nodo Storage (CEPH Bricks)

Min 48core fisici
RAM >512GB DDR4 o DDR5
Almeno 360 TBL HDD + 12TBL SSD



Accessori

Switch IB, Switch ETH
Cavi IB, Cavi ETH
Transceiver vari
Assistenza 3+2

HPC Bubbles

- Accordo Quadro Nazionale
 - Listino prezzi per nodi + accessori
 - 2 anni di validità
 - Lotto1
 - CPU, GPU, FPGA
 - Lotto2
 - Storage
 - Sedi Coinvolte: CNAF, BARI, MI-BI, PI, TO, LNGS, NA, RM1, PD/LNL
- **Stato gara**
 - **Bando Scaduto**
 - **4 offerte**
 - **Commissione in attesa di nomina ufficiale per definire data apertura buste tecniche**

Quantità stimate – non definitive

Sito	Nodi CPU	Nodi GPU	Nodi FPGA	Storage (PBN EC)
CNAF	16	14	2	4
CNAF-ISO27001	8	8	0	2
BA	7	3	0	3
MI-BI	0	0	4	0
PI	8	0	0	0
TO	6	4	0	0
LNGS	0	4	0	0
NA	8	0	2	0
RM1	8	0	0	0
PD/LNL	6	6	0	0
TOTAL	60	35	6	9

Gara Nazionale Storage - 1

ACQUISIZIONI LOTTO1

SEDE	NOME SISTEMA	CAPACITÀ MINIMA SISTEMA (TBL)
INFN Sezione di Bari	ST-BA-1	3.789,5
	ST-BA-2	3.789,5
INFN Sezione di Catania	ST-CT	1.625,0
INFN Laboratori Nazionali di Legnaro	ST-LNL-1	1.020,5
	ST-LNL-2	1.358,5
INFN Sezione di Torino	ST-TO	2.379,0
TOTALE		13.962,0

- La gara riguarda un insieme di sistemi di storage di almeno **26598 TB Lordi**
 - suddivisi in **due lotti** rispettivamente di **13962 TBL** e **12636 TBL**
 - 8 sedi coinvolte: BA, CT, TO, LNL, MI, NA, RM1, LNF
- **Lotto 1 – ENCLOSURE DENSIS : lunghezza massima apparati 990mm**
 - Unità con almeno 2 controller FC 4x32Gbs
 - Compatibilità con switch Brocade
 - Minimo 12 dischi per RU
 - 14TB o 16TB per disco
 - Inclusi disk server in numero adeguato
 - Almeno 5 anni di garanzia e assistenza tecnica NBD on-site
- 1.057.377,05 + IVA base di gara

Gara Nazionale Storage - 2

ACQUISIZIONI LOTTO2

SEDE	NOME CONVENZIONALE SEDE	NOME SISTEMA	CAPACITÀ MINIMA SISTEMA (TBL)
INFN Laboratori Nazionali di Frascati	INFN-LNF	ST-LNF	2.379,0
INFN Sezione di Milano	INFN-MI	ST-MI	2.223,0
INFN Sezione di Napoli	INFN-NA	ST-NA	5.655,0
Sezione INFN Roma 1	INFN-RM1	ST-RM1-1	754,0
		ST-RM1-2	1.625,0
TOTALE			12.636,0

- **Lotto 2: lunghezza massima apparati 850mm**
 - Sono inclusi almeno 5 anni di garanzia + assistenza tecnica NBD on-site

- **Lotto 2 – ENCLOSURE NON-DENSI : lunghezza massima apparati 850mm**
 - Unità con almeno 2 controller FC 4x32Gbs
 - Compatibilità con switch Brocade
 - Minimo 6 dischi per RU
 - 14TB
 - Inclusi disk server in numero adeguato
 - Almeno 5 anni di garanzia e assistenza tecnica NBD on-site
 - 965.635,41 + IVA base di gara

Gara Nazionale Storage - 3

ACQUISIZIONI LOTTO2

SEDE	NOME CONVENZIONALE SEDE	NOME SISTEMA	CAPACITÀ MINIMA SISTEMA (TBL)
INFN Laboratori Nazionali di Frascati	INFN-LNF	ST-LNF	2.379,0
INFN Sezione di Milano	INFN-MI	ST-MI	2.223,0
INFN Sezione di Napoli	INFN-NA	ST-NA	5.655,0
Sezione INFN Roma 1	INFN-RM1	ST-RM1-1	754,0
		ST-RM1-2	1.625,0
TOTALE			12.636,0

STATO GARA

• 8 ditte hanno presentato offerta (6 offerte per lotto1 e 5 offerte per lotto2): passano l'analisi tecnica 6 ditte per il lotto1 e 1 ditta per il lotto2.

• Aperte buste economiche, ribassi tra il 27 e il 30%.

• Lotto1 29,95% di capacità di disco in più, 7 anni garanzia. (65 E/TBN IVA inclusa)

• Lotto 2 30,01% di capacità di disco in più, 7 anni garanzia (65 E/TBN IVA inclusa)

• In corso verifica anomalia dell'offerta



Gara nazionale concentratori - 1

- Le sedi interessate dall'acquisto nazionale sono LNGS, LNL, PD, Roma1
 - LNGS 2 apparati
 - LNL 1 apparato
 - PD 1 apparato
 - Roma1 1 apparato
- Si acquisteranno quindi 5 apparati tutti uguali tra loro
- Per ciascun apparato:
 - 5 bretelle MTP/MPO-12 F/F metodo B di 10 metri in fibra ottica multimodale OM4
 - 5 breakout cable MTP/MPO-12 to LC Duplex (4xLC) in fibra ottica multimodale OM4 di 10 metri (Da usare per "Splittare" le porte 40Gb e 100 Gb rispettivamente in 4x10Gb e 4x25 Gb)
- Si acquistano transceiver originali e nuovi
 - 42 transceiver 100G BASE-SR4
 - 8 transceiver 100G BASE-LR4
 - 4 transceiver 100G BASE-ER4
 - 10 transceiver 40G BiDi (Su singola coppia di fibre multimodali)
 - 12 transceiver 40G BASE SR4

STATO procedura

- **Espletata**
- **In preparazione documenti per AC con quote per le varie sedi**
- **Ribasso medio 3%**
- **1 sola ditta ha presentato offerte**
 - **la stessa degli apparati attivi per il Tecnopolo**

Gara nazionale concentratori - 2

DETTAGLIO PER SEDE									
Nome Sede	Indirizzo Sede	N. apparati	Air Flow	N. Transceiver 40G SR4	N. Transceiver 40G BiDi	N. Transceiver 100G SR4	N. Transceiver 100G LR4	N. Transceiver 100G ER4	Referente Sede
INFN Laboratori Nazionali di Legnaro	Viale dell'Università, 2 - 35020, Legnaro (PD) - Italia	1	Front to Back	0	0	12	4	2	Michele Gulmini
INFN Sezione di Padova	Via Vesalio 11, 35131, Padova	1	Back to Front	7	0	8	0	2	Massimo Sgaravatto
INFN Laboratori del Gran Sasso	Via Acitelli 22, 47100, Assergi, (AQ)	2	Front to Back	4	0	16	2	0	Sandra Parlati
INFN Sezione di Roma1	C/o Dipartimento di Fisica - Università Sapienza, Piazzale Aldo Moro 2, 00185, Roma	1	Back to Front	1	10	6	2	0	Alessandro De Salvo
Totali		5		12	10	42	8	4	

Gara nazionale concentratori - 3

- Ogni apparato dovrà avere le seguenti caratteristiche
 - Switch 1 RU
 - 36 porte di cui 32 porte 100 Gb Ethernet che possano ospitare transceiver di tipo 100G QSFP28 ed in aggiunta, almeno 4 porte 400Gb Ethernet (QSFP-DD); Nel calcolo delle porte a 100Gb si ammettono anche porte a 400Gb che possano ospitare nativamente transceiver a 100Gb (QSFP-28).
 - Le porte a 100 Gb devono poter essere configurate e/o frazionate anche tramite cavi di breakout a velocità inferiori (10/25/40 Gb).
 - Le porte a 400 Gb devono potere essere configurate e/o frazionate a velocità inferiori.
 - Dall'apparato anche tramite l'utilizzo di cavi di breakout dovrà essere possibile ricavare almeno: 36 connessioni a 40Gb, 64 connessioni a 25Gb, 64 connessioni a 10Gb
 - Alimentazione ridondata
 - Capacità di switching $\geq 4\text{Tbps}$
 - Deep buffer $\geq 8\text{GB}$

Gara nazionale concentratori - 4

- Ogni apparato dovrà supportare le seguenti feature:
 - Routing IPv4 ed IPv6 a “wire speed”
 - Jumbo Frame (fino ad una MTU di almeno 9216 Bytes).
 - ACL di sicurezza complesse con almeno 8000 entry complessive in IPv4/IPv6 applicabili sia in ingresso che in uscita contemporaneamente, a wire speed (in ASIC senza interessare le CPU dell'apparato).
 - Configurazione via linea di comando (CLI) e dovranno supportare anche strumenti di automazione, tra cui Ansible.
- Gli apparati proposti dovranno avere le seguenti funzionalità e supportare a pieno i seguenti protocolli:
 - IPv4 ed IPv6 (Compresa la funzionalità di DHCPv6 Relay)
 - SNMP v2 o versioni superiori; 802.1Q, ; VXLAN (Routing e Bridging); EVPN; 802.1ad ed LACP; Spanning Tree (RSTP,PVST); DHCP Relay; OSPF, BGP, Equal Cost Multipath, MPLS; VRRP e/o HSRP; VARP (Virtual ARP); ESI (Ethernet Segment Identifier); Policy based routing; VRF (Con. VRF Leaking); NetFlow o protocollo equivalente per l'analisi dei flussi; Span Port

Gara nazionale concentratori - 5

- Assistenza in garanzia NBD on-site per un periodo minimo di 5 anni su tutto il materiale fornito
- Consegna presso le quattro sedi/strutture entro 120 giorni naturali e consecutivi dalla data di sottoscrizione del contratto
- Formazione e affiancamento all'avvio operativo delle apparecchiature fornite dovrà essere svolta entro e non oltre 30 giorni naturali e consecutivi decorrenti dalla data di installazione degli apparati presso le sedi/strutture della Stazione Appaltante

CPU Adesione convenzione CONSIP

• Tecnologie Server 4 – Lotto 8 - TWINSQUARE DELL POWEREDGE C6400, 4XC6525

- 2xAMD7413 24C/48T per nodo
- 512GB RAM per nodo
- 2x1TB SSD + 2x4TB SSD o 2x1TB SSD + 4x2.4 TB HDD 10k
- Garanzia 3+2

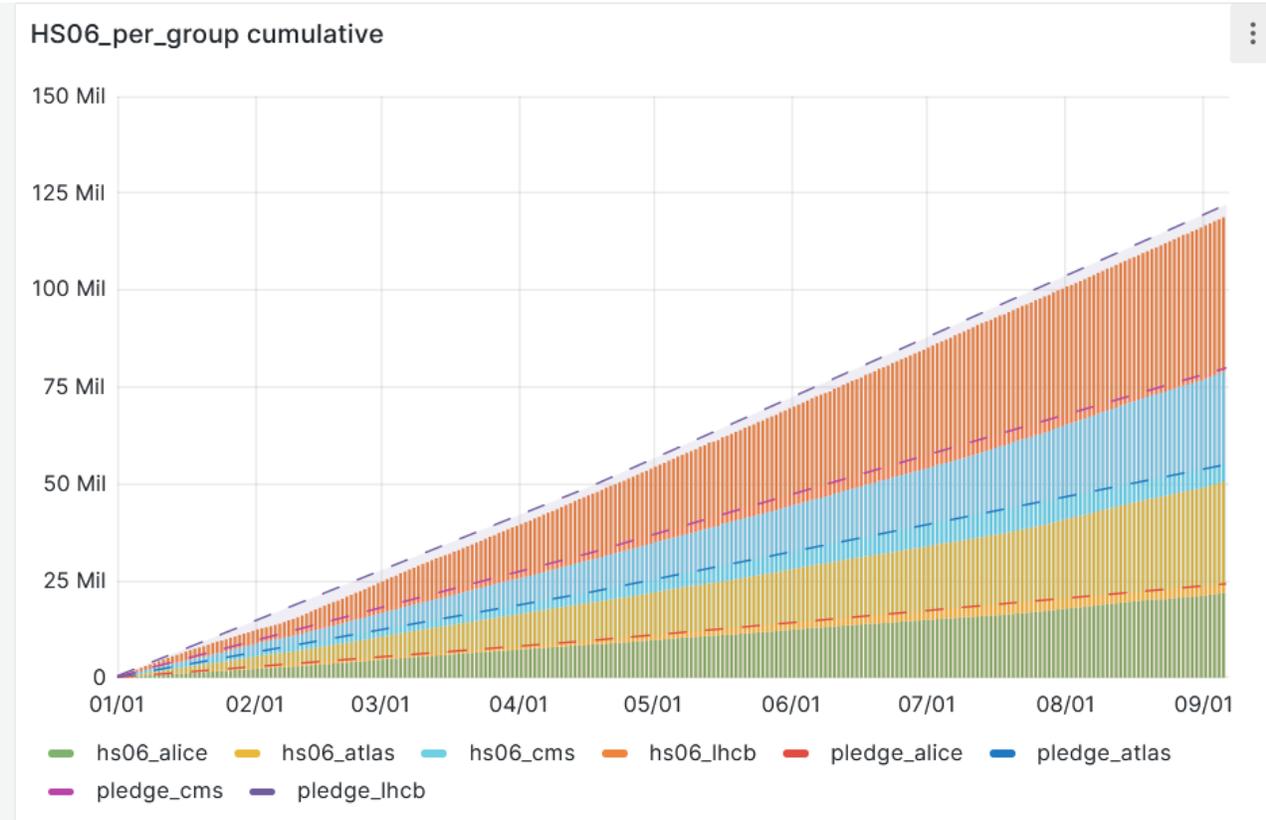
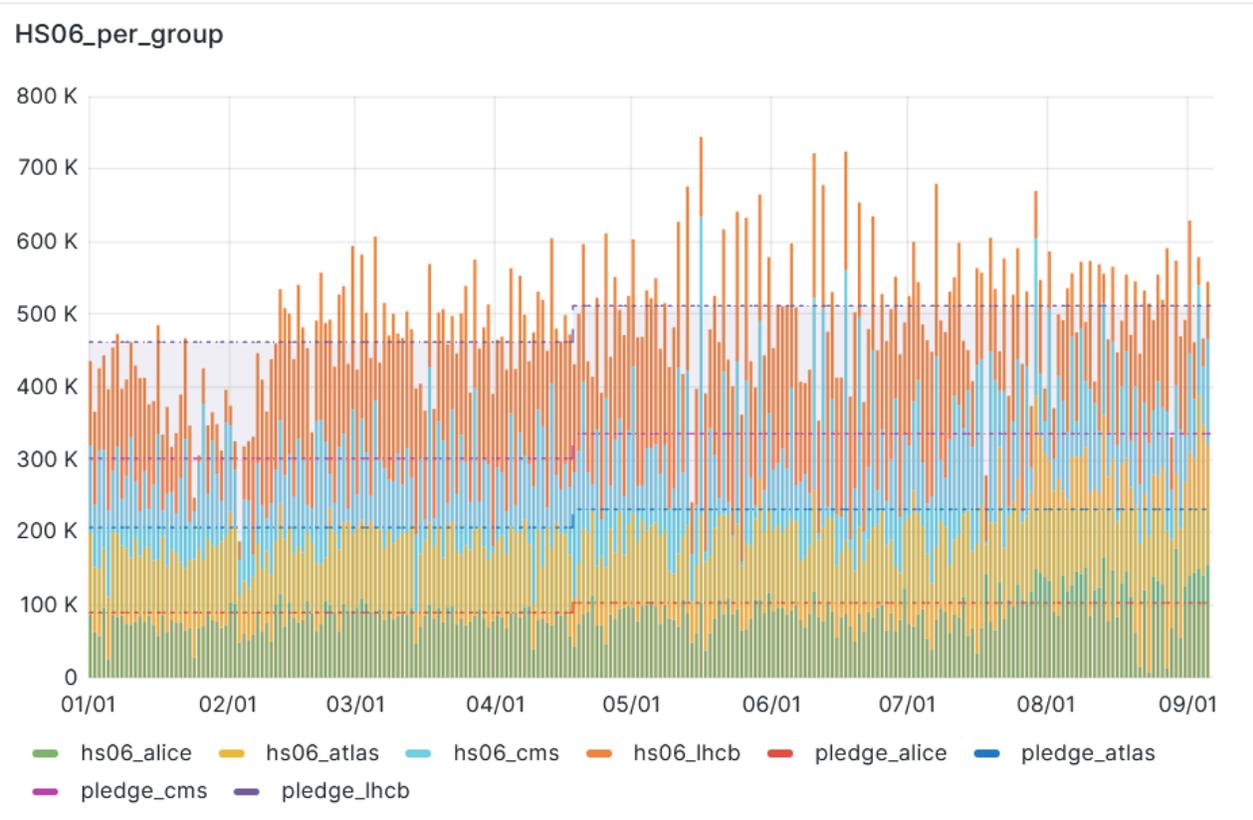
SEDE	Numero sistemi
BARI	19
NAPOLI	19
CATANIA	7
ROMA1	8
FRASCATI	8
TORINO	9
MILANO	8
PISA	9
LEGNARO	8

Acquisizioni molto simili sui siti a meno di piccolo cambiamenti di configurazione in particolare sui dischi

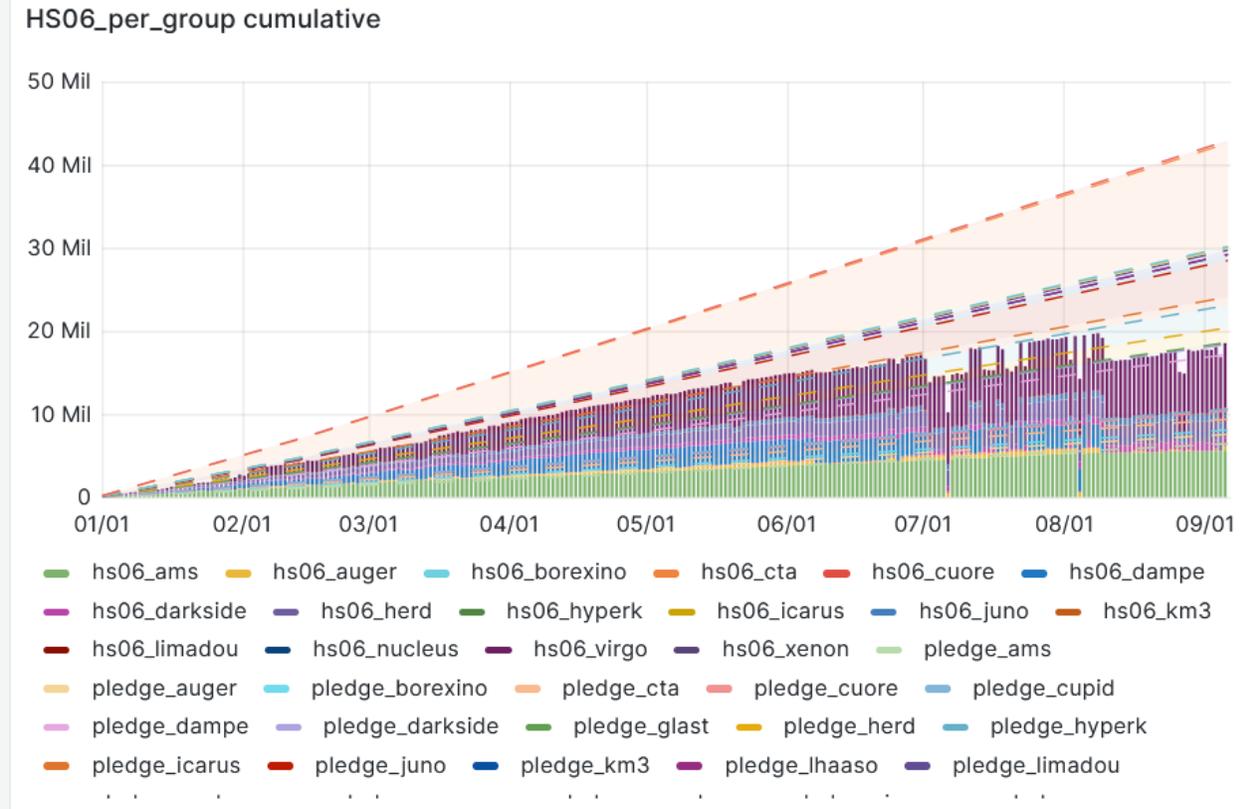
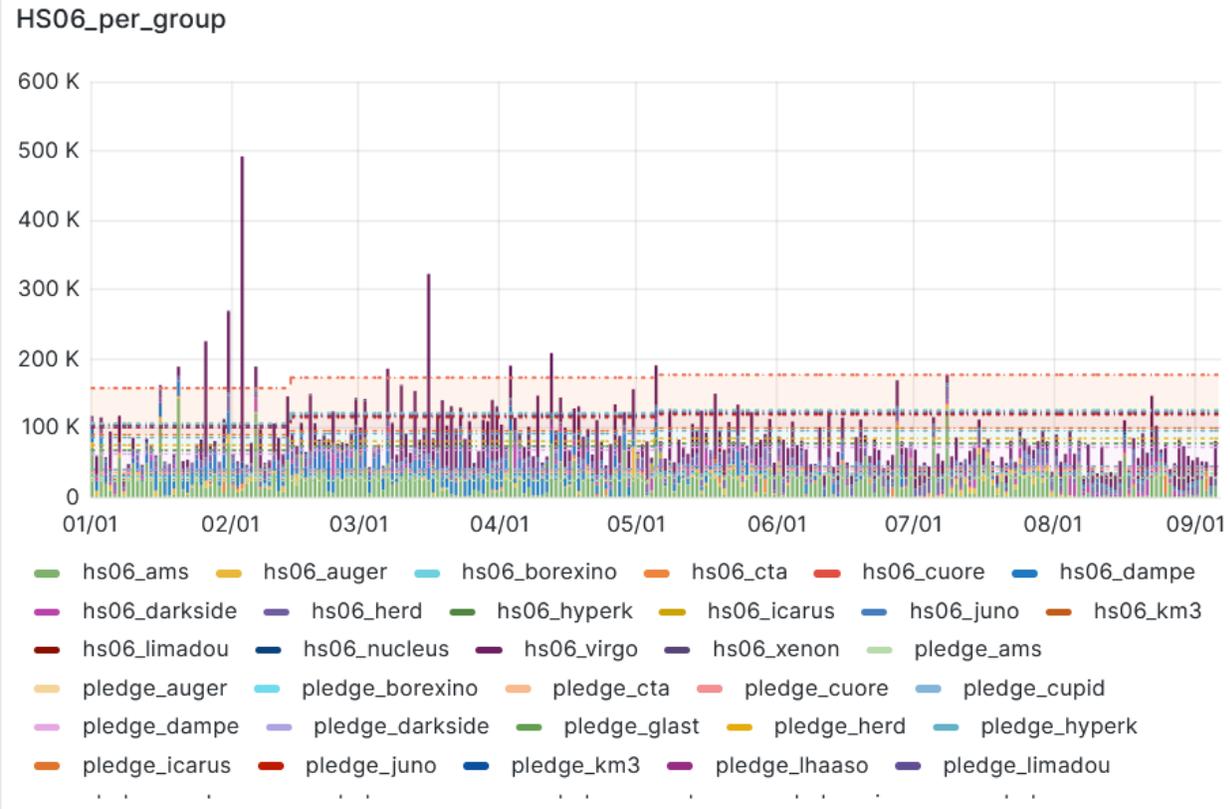


BACKUP

CPU Usage LHC 2023

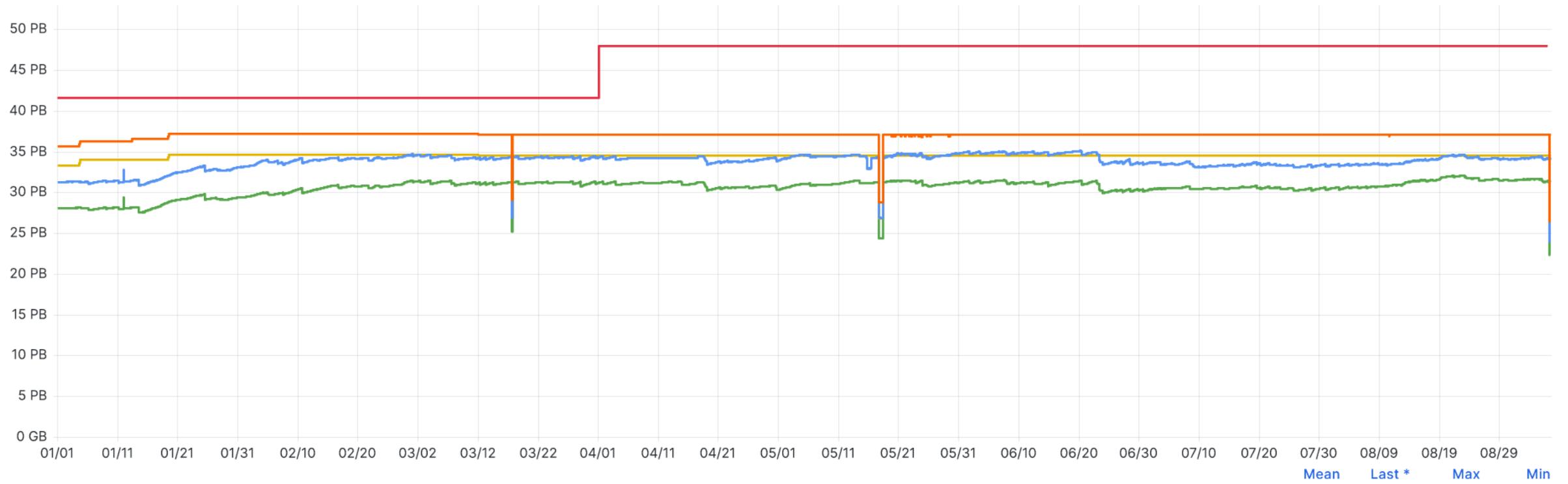


CPU Usage CSN2 2023



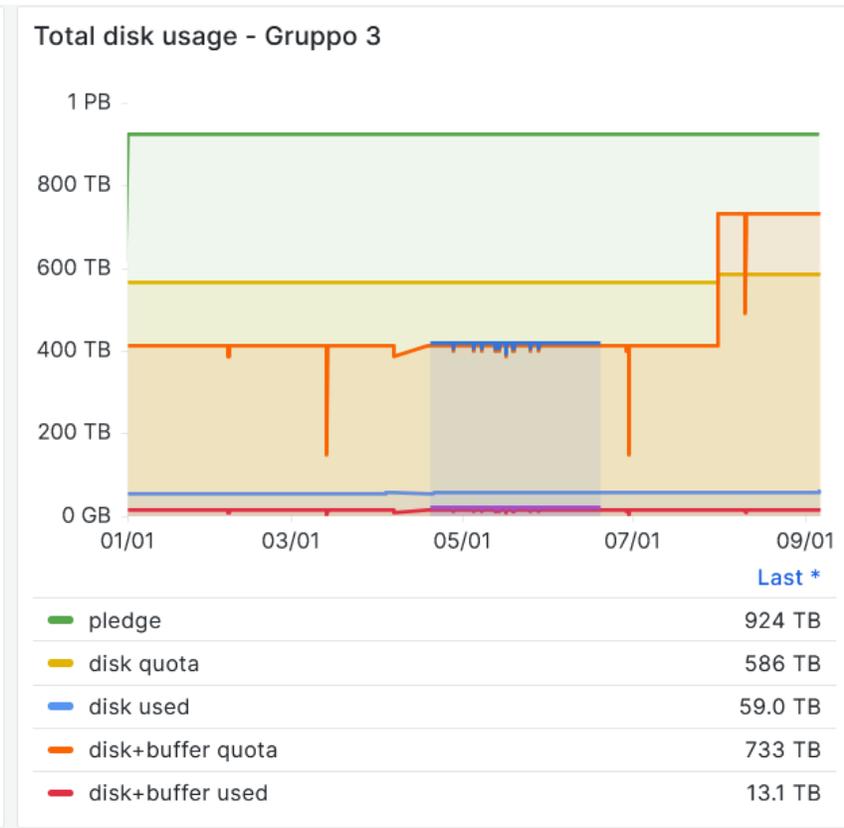
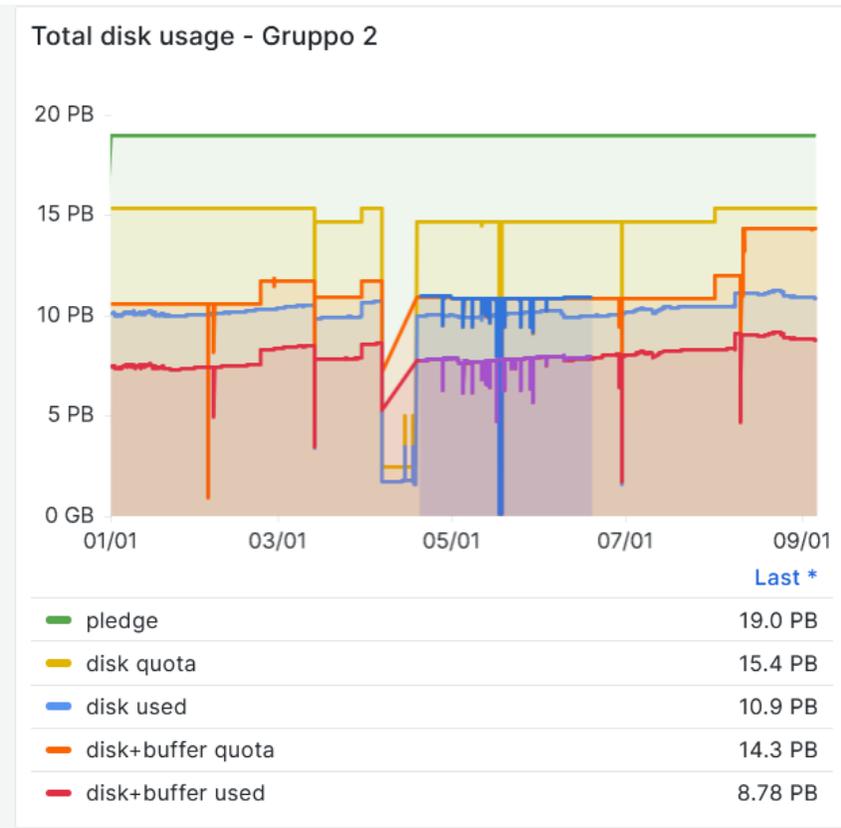
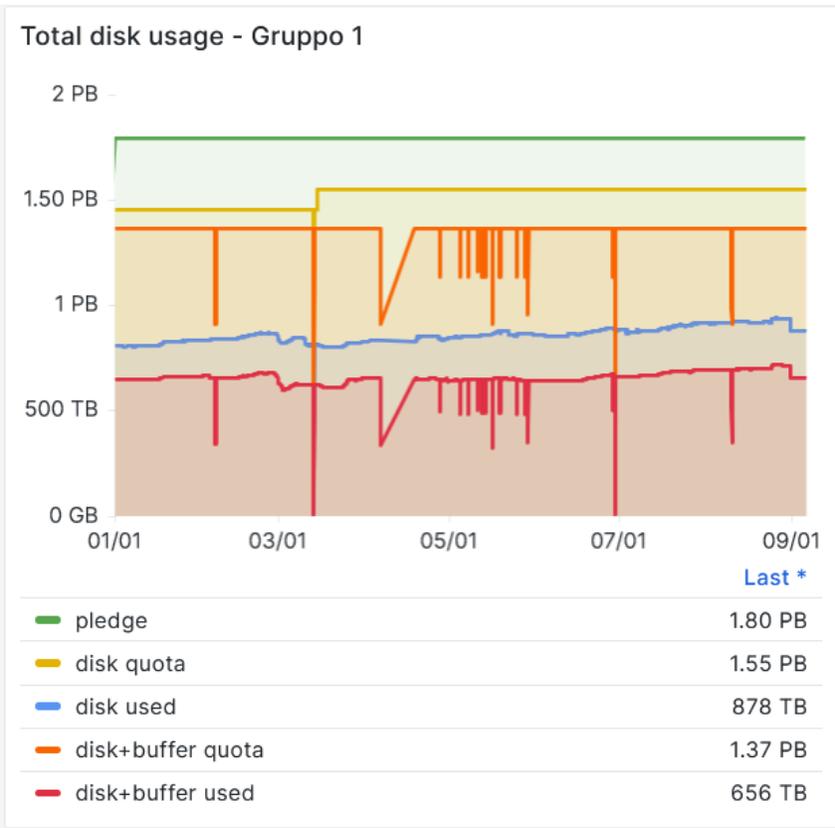
Global Disk Usage LHC 2023

Total disk usage (LHC)

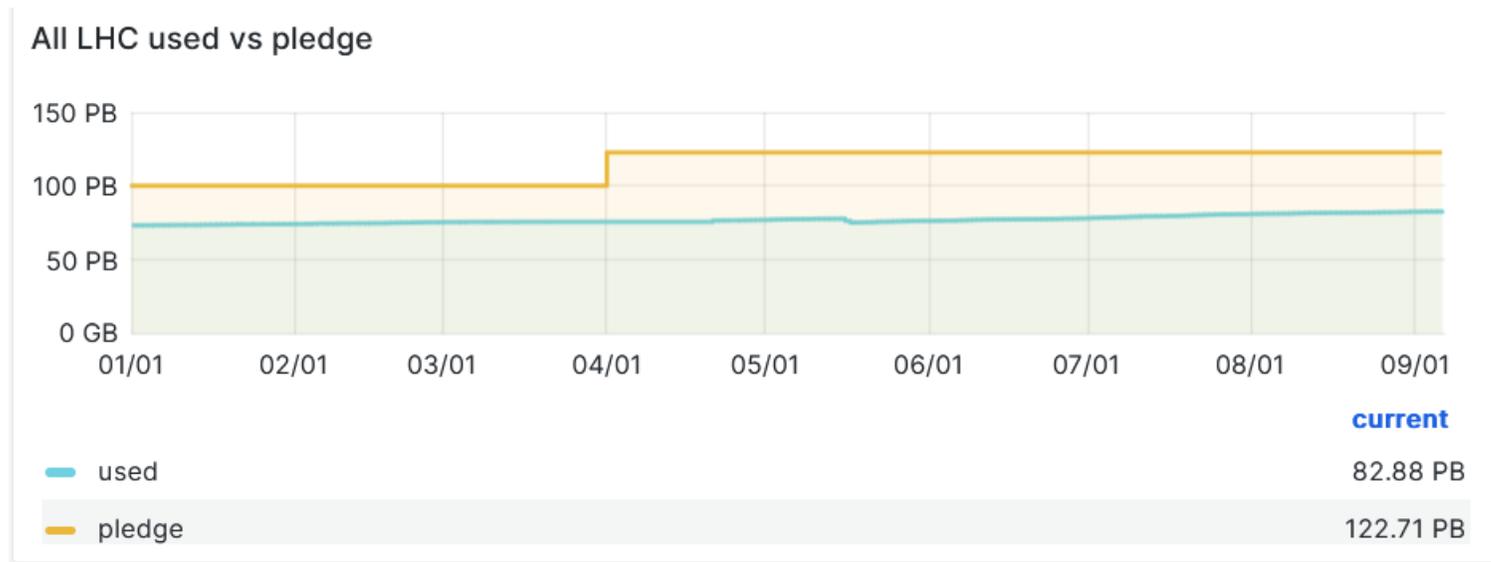


	Mean	Last *	Max	Min
disk used	22.2 PB	31.4 PB	32.1 PB	1.41 PB
disk quota	26.0 PB	34.5 PB	35.5 PB	1.47 PB
disk+buffer used	25.6 PB	34.3 PB	35.1 PB	293 TB
disk+buffer quota	26.6 PB	37.1 PB	37.2 PB	3.16 MB
pledge	45.6 PB	48.0 PB	48.0 PB	41.6 PB

Global Disk Usage NOLHC 2023

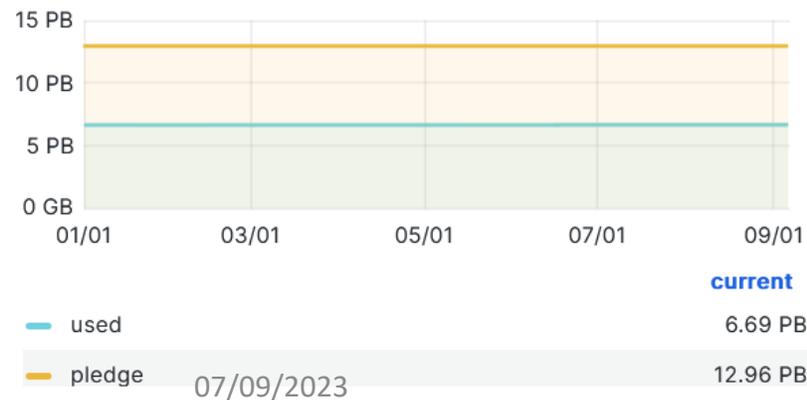


Tape Usage LHC

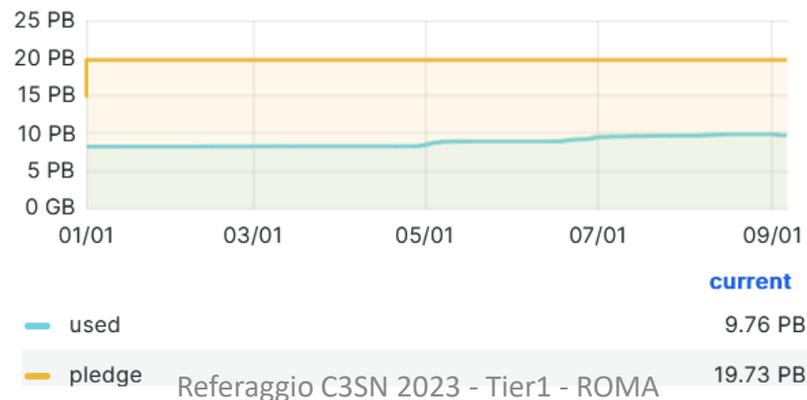


Tape Usage NoLHC

Total used vs pledge - Gruppo 1



Total used vs pledge - Gruppo 2



Total used vs pledge - Gruppo 3

