

Preventivi CMS 2024

Daniele Spiga INFN-PG
On behalf of CMS-Calcolo Italia

La Sapienza 7.09.2023

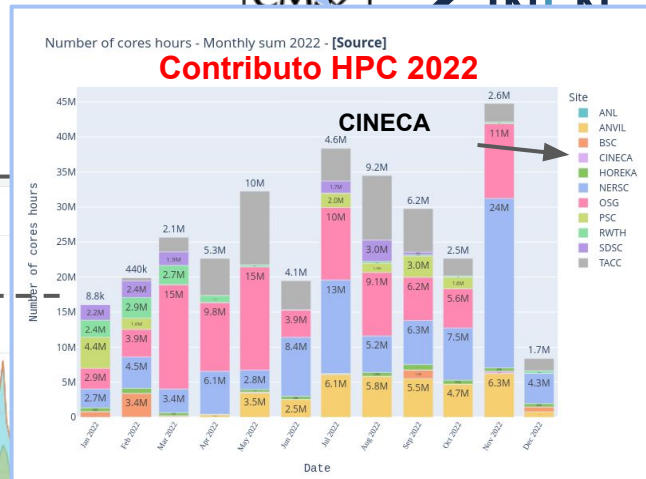
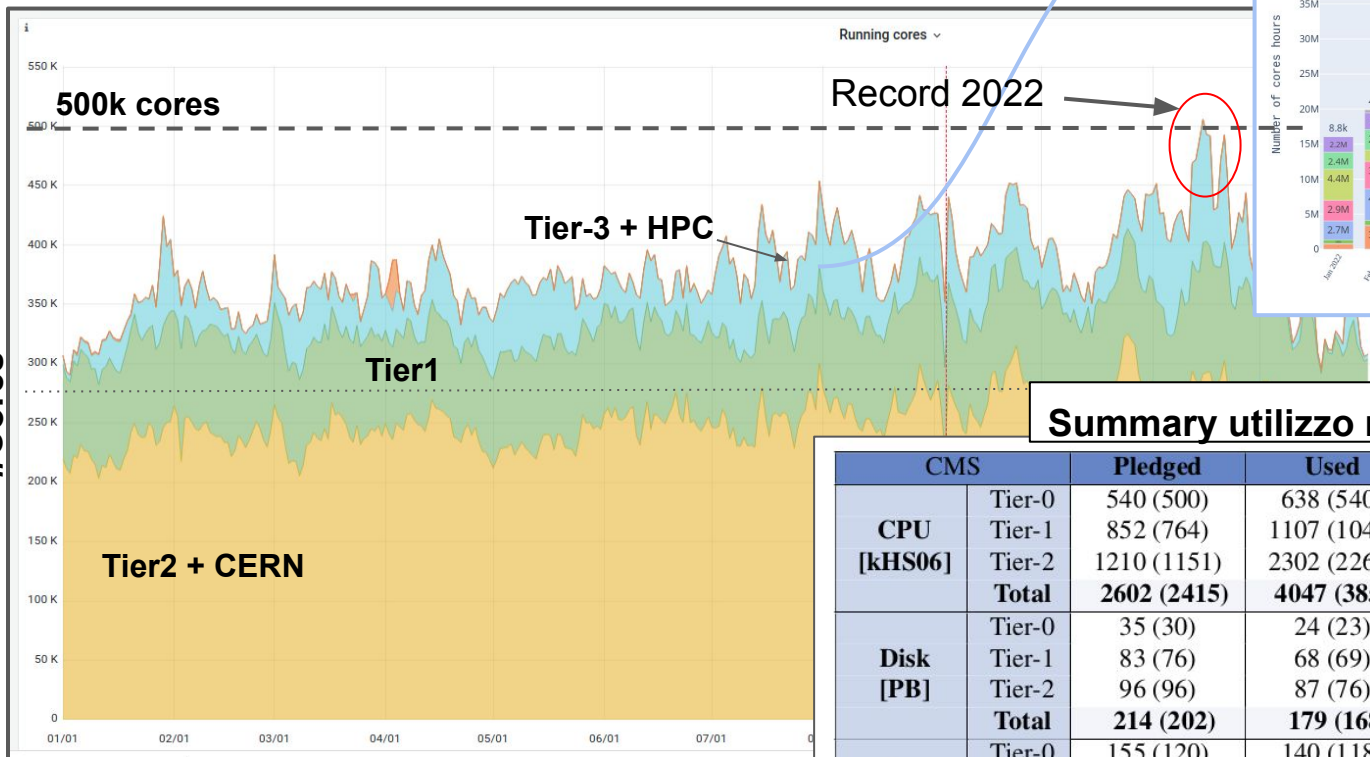
Agenda



- Breve(issimo) report sull'uso delle risorse 2022
- Richieste per il 2024
 - Andamento storico e proiezioni
 - Dismissioni HW (e possibili scenari di mitigazione)
 - Nuove richieste
- Update Tier1 a CMS (JINR e candidati T1)

Risorse calcolo@CMS: contributi e utilizzo

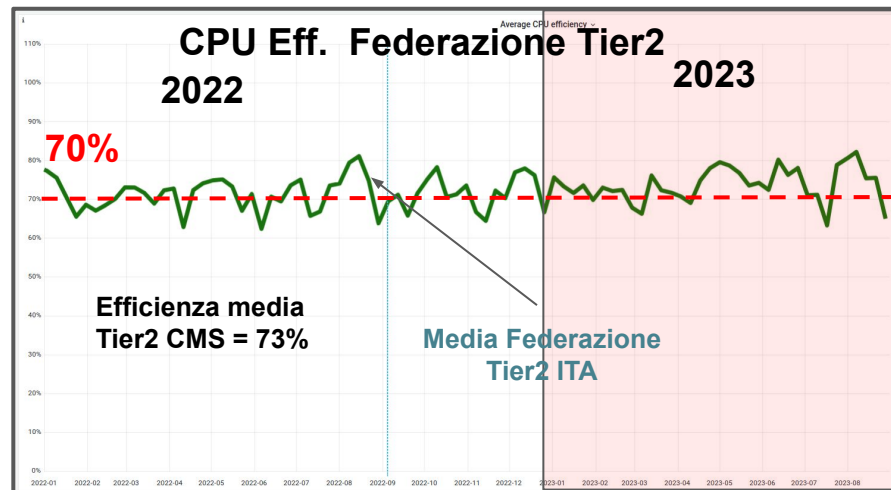
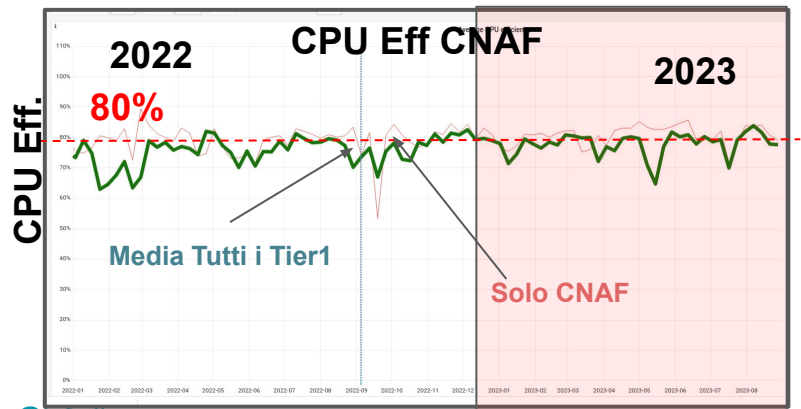
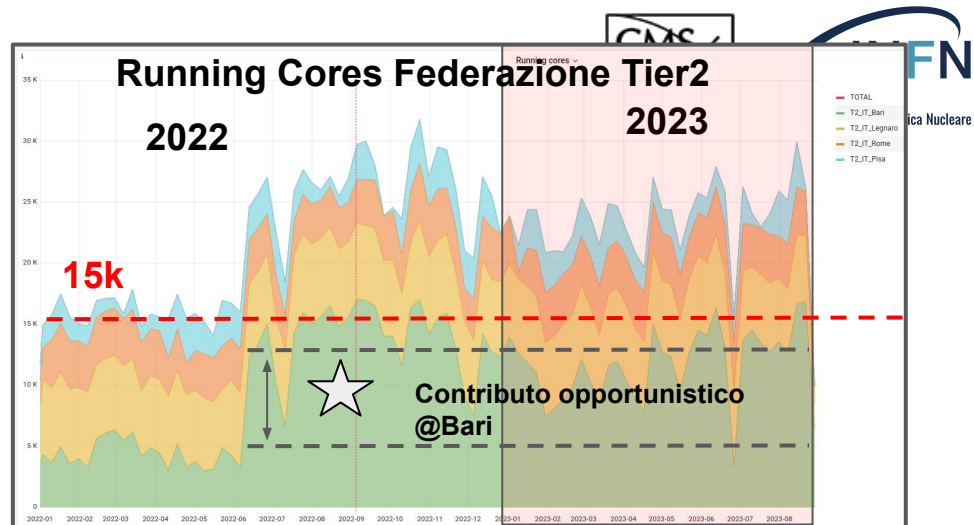
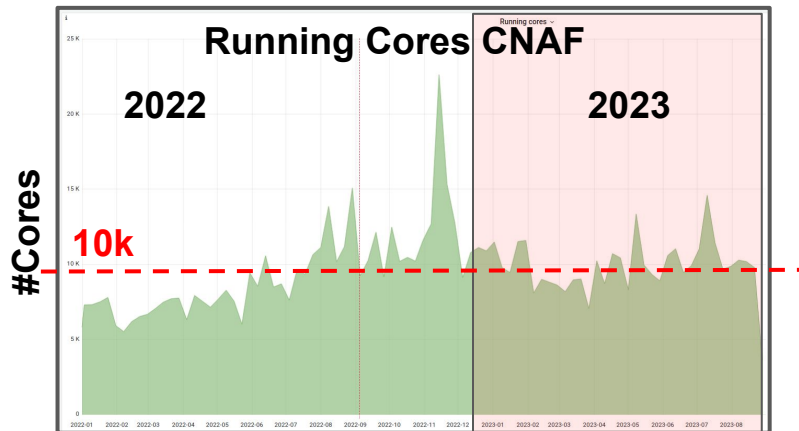
#Cores



Summary utilizzo risorse 2022

CMS		Pledged	Used	Used/Pledged	Avg. CPU Eff
CPU [kHS06]	Tier-0	540 (500)	638 (540)	118% (108%)	74% (71%)
	Tier-1	852 (764)	1107 (1047)	130% (137%)	83% (80%)
	Tier-2	1210 (1151)	2302 (2263)	190% (197%)	73% (70%)
	Total	2602 (2415)	4047 (3850)	156% (159%)	76% (73%)
Disk [PB]	Tier-0	35 (30)	24 (23)	68% (76%)	
	Tier-1	83 (76)	68 (69)	82% (91%)	
	Tier-2	96 (96)	87 (76)	91% (79%)	
	Total	214 (202)	179 (168)	84% (83%)	
Tape [PB]	Tier-0	155 (120)	140 (118)	90% (98%)	
	Tier-1	255 (219)	179 (211)	70% (96%)	
	Total	410 (339)	319 (329)	78% (97%)	

I siti Italiani

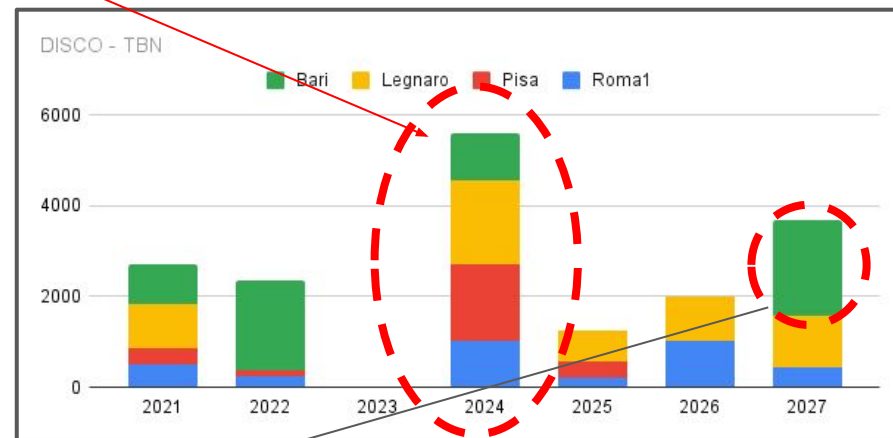
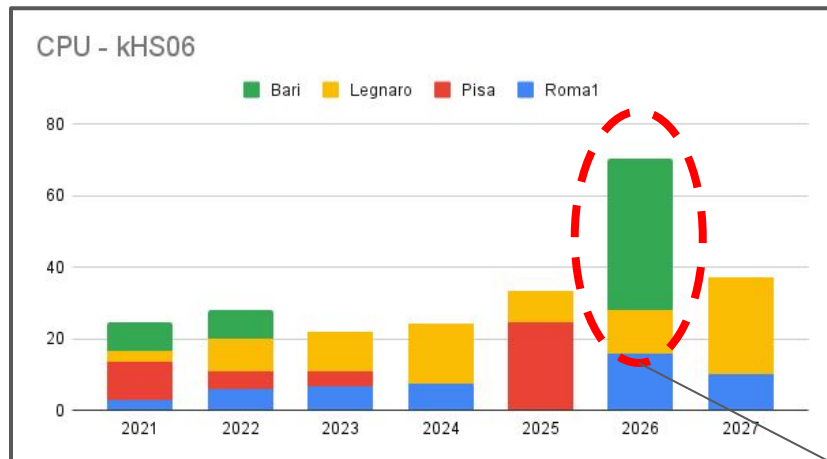


Dismissioni HW Tier2: Summary e proiezioni



Picco di dismissione effetto gara nazionale 2018/19 (Huawei)

**Proiezione
2027 non
accurata**



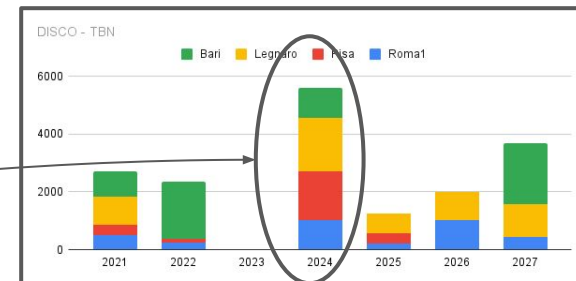
**Effetto bulk injection a Bari PON
IBISCO**

Dismissione HW Tier2



Per la monetizzazione si
è fatto riferimento a
questo preziario:

Tape 10 euro / TB
Disco 120 euro / TBN
CPU 10 euro / HS06



	Disco (TBN)	CPU (kHS06)	\$Disco	\$CPU
Bari	1035	0	124200	0
Pisa	1708	0	204960	0
Legnaro	1840	16.7	220800	167000
Roma1	1008	0	120960	0
TOT	5591	16.7	670920	167000

Possibile mitigazione su dismissione Disco



Il grosso della dismissione 2024 è relativo a sistemi Huawei (della gara nazionale 2018/19). Nessuno di noi ritiene economicamente conveniente effettuare un'estensione di garanzia per questo HW.

Esistono due possibili maniglie

- 1) **Roma ha 900 TB di storage DELL** in dismissione per il quale avrebbe senso estendere la manutenzione
 - Costo 13.5 kEuro per 1 anno (2025) e 24.5 kEuro per 2 anni (2026).
 - Attenzione che nel 2026 a Roma avremo >1PB che andrà sostituito
- 2) **Pisa ha circa 700 TB su DDN** per il quale avrebbe senso estendere la manutenzione
 - Prezzo stimato è 18k per arrivare al 2025
 - Inoltre questo sistema ha ancora spazio per aggiungere ulteriori dischi

Dismissioni Disco Tier2: Scenario #2



Si riducono i contributi Pisa/Roma sulla sola parte Storage, diventando questo

	Disco (TBN)	CPU (kHS06)	\$Disco	\$CPU
Bari	1035	0	124200	0
Pisa	1708	0	204960	0
Legnaro	1840	16.7	220800	167000
Roma1	1008	0	120960	0
TOT	5591	16.7	670920	167000

	Disco (TBN)	CPU (kHS06)	\$Disco	\$CPU	OverHead Server	OverHead Networking
Bari	1035	0	124200	0	8694	6210
Pisa	1000	0	120000	0	8400	6000
Legnaro	1840	16.7	220800	167000	27146	21060
Roma1	100	0	12000	0	840	600
TOT	3975	16.7	477000	167000	45080	33870

Richieste CMS 2024



CMS		'23 Approved Request - Spring '22	'24 Preliminary Request - Fall '22	'24 Final Request Spring '23	Increase wrt '23 Abs. Perc.	
CPU [kHS06]	Tier-0	720	750	980	260	36%
	Tier-1	800	860	☆ 930	130	16%
	Tier-2	1,350	1,500	☆ 1,600	250	19%
	Total	2,870	3,110	3,510	640	22%
Disk [PB]	Tier-0	45	52	54	9	20%
	Tier-1	98	108	☆ 122	24	24%
	Tier-2	117	130	☆ 149	32	27%
	Total	260	290	325	65	25%
Tape [PB]	Tier-0	228	293	320	92	40%
	Tier-1	316	370	☆ 380	64	20%
	Total	544	663	700	156	29%

WLCG RRB 25.04.2023
ha approvato le richieste
riportate in questa tabella

**Consideriamo sempre il
contributo italiano pari al
13% anche se siamo
cresciuti al 14%**

- (vedi MoU Backup)

Parameter	Appr. 2023	New 2023	2024
LHC			
LHC Energy pp [TeV]	14*		13.6
Average pileup	50	62 (peak at 65)	
Integrated luminosity / year [fb^{-1}]	100	90	110
Livetime pp / year [$\text{s}/10^6$]	6	4.2	5.2
Livetime HI / year [$\text{s}/10^6$]	1.2	1.7	1.7
Heavy Ion run type		Pb-Pb	
CMS-Specific			
Prompt HLT Rate [kHz]	1.7		2.6
Parked HLT Rate [kHz]		3	
Scouting Rate [kHz]		30	
Run 3 MC events / year in billions	29	45	53
Phase-2 MC events / year in billions		0.5	

Parametri utilizzati per la definizione delle
richieste sopra.

- Parametri della Macchina
- Parametri specifici di CMS

Nuove richieste 2024



Distribuzione flat di crescita CPU e Storage sui 4 siti della Federazione

	Disco (TBN)	CPU (kHS06)	\$Disco	\$CPU
Bari	1040.0	8.125	124796	81250
Pisa	1040.0	8.125	124796	81250
Legnaro	1040.0	8.125	124796	81250
Roma1	1040.0	8.125	124796	81250
TOT	4159.9	32.5	499182	325000

Gran Totale CMS per 2024



	TOT (euro)	disco TBN	disco Eur	CPU kHS06	CPU Eur	Rete Eur	Server Eur
Bari	370687.46	2075.0	248996	8.1	81250	17324.775	23117.185
Pisa	461138.66	2748.0	329756	8.1	81250	21362.775	28770.385
Legnaro	667589.46	2880.0	345596	24.8	248250	32174.775	41569.185
Roma1	367058.66	2048.0	245756	8.1	81250	17162.775	22890.385
TOT	1866474.24	9750.85	1170102.00	49.2	492000.00	88025.10	116347.14

L'impatto della mitigazione sul totale è del **11.5%** circa

	TOT (euro)	disco TBN	disco Eur	CPU kHS06	CPU Eur	Rete Eur	Server Eur
Bari	370687.46	2075.0	248996	8.1	81250	17324.775	23117.185
Pisa	365983.46	2040.0	244796	8.1	81250	17114.775	22823.185
Legnaro	667589.46	2880.0	345596	24.8	248250	32174.775	41569.185
Roma1	245023.46	1140.0	136796	8.1	81250	11714.775	15263.185
TOT	1649283.84	8134.85	976182.00	49.2	492000.00	78329.10	102772.74

NOTA: in questo scenario per Pisa e Roma dobbiamo aggiungere i costi dell'estensione (+10k Pisa +13k Roma per 1 anno)

- E questo andrebbe come "anticipo di bilancio"

Nuove Richieste per il Tier1



È Stato considerato il 13% e NON è stato incluso
nessun “effetto JINR”

RICHIESTE 2024				
T1	pledge 2023	pledge 2024	Incremento	Eur Delta
CPU (kHS06)	104	120.9	16.9	169000
DISK (TBN)	12740	15680	2940	352800
TAPE (TB)	41080	49400	8320	83200
				605000

Situazione Tier1 a CMS (JINR)



Per il 2023 praticamente tutte le FA
(**Grazie!**) hanno aiutato CMS con Extra
Pledge per il TAPE

Per il futuro CMS sta lavorando alla
ricerca di un nuovo(i) Tier1. Attualmente
ci sono due candidati (Serbia e Polonia)

Tier-1 site	2023 tape pledge [PB]	% of Request	Beyond-pledge tape [PB]
KIT	31.6	10.0	6.3
PIC	12.64	4.0	1.0
IN2P3	32.46	10.3	
CNAF	41.01	13.0	5.5
RAL	24.42	7.7	2.0
FNAL	126.4	40.0	22.0
Total	268.6	85	36.8

Nota: non
ancora
“installato”

Proposal is to construct Serbia CERN Tier-1 center in 2 stages

- Stage 1:
 - Disk storage 10-20 PB
 - Tape: 40 PB
 - CPU: 11.5 k cores
 - LHCOPN @ 1-2x 100 Gbps
 - ETA Q1 2024
- Stage 2:
 - Disk storage 40 PB:
 - Tape: 80 PB
 - CPU/ GPU: 25 k cores
 - LHCOPN @ 4-10x 100 Gbps
 - ETA Q4 2025

Timeline:

- Q1/2024 Procurement and bring-up equipment and services
- Q1/2024-Q1/2025 Go through probation period of 1 year
- Q1/2025 Become fully operational recognized CERN Tier-1
- Take advantage of excellent facility of Serbian State Data Center.
 - Connectivity, Power, Cooling (2N redundancy)
- Train experts, bring know-how and participate in cutting-edge research in the field of Distributed Computing.
 - large pool of human resources in IT in the country
- Put in use local infrastructure for Science and Technology.

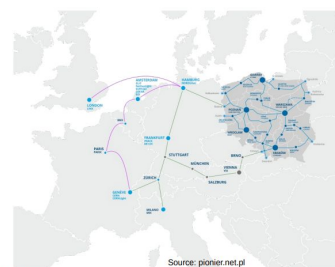
7/21/23 Serbia CERN Tier-1 Proposal, V. Nekovic

NCBJ-CIS WLCG Proto-Tier 1 Network resources

- 100 Gbps link to PIONIER (academic internet, Geant)
 - 20 Gbps dedicated channel to LHCONE
 - Full speed achieved (2022 Data Challenge)
 - 20 Gbps dedicated channel to LHCOPN
 - Backup path to be established
- Two independent 8 Gbps links to internet (Exatel)
 - Provides around 5.6 Gbps to CERN (2022 Data Challenge)
 - Backup links for WLCG activities
- 2 x 40Gbps internal backbone
- 2 x 10Gbps external firewall links (upgrade in plans)

Milestones

- 1) Dedicated LHCOPN VLAN (end of April 2023) (done)
- 2) Support of IPv6 on the academic internet links (ongoing)



Source: pionier.net.pl

Organigramma Offline & Computing



Offline Software and Computing - June 2023

Coordinators

D. Piparo, J. Letts

Core Software

S. Muzaffar, M. Kortelainen

Simulation

V. Ivantchenko, S. Bein

Monitoring & Analytics

F. Legger, B. K. Jashal

Analysis Infra. & Support

S. Belforte, K. Ellis

Generators *

S. Bhattacharya, M. Lu

Computing Resources Board

J. Hernandez, K. Bloom

Computing Operations

A. Wightman, D. Kovalskyi

Workload/Data Mgt. Devel.

K. Lannon, D. Ciangottini

Offline Release Planning

A. Perrotta, S. Rappoccio

Upgrade R&D and TDR

D. Elvira, F. Ferri

Machine Learning *

G. Kasieczka, J. Ngadiuba

Dyn. Res. Provisioning

D. Spiga, C. Wissing

Reconstruction

C. Caputo, M. Nguyen

Upgrade Software

P. Srimanobhas, A. Di Florio

Web Services & Security

A. Pfeiffer, M. Imran

L1 Software **

E. PalenciaCortezon, A. Loeliger

Facility Services

G. Bagliesi, S. Lammel

Resource Management

J. Flix, D. Lange

Submission Infra.

A. Perez-Calero, M. Mascheroni

DPOA ***

K. Lassila-Perini, J. Hogan

* Joint with Physics / ** Joint with L1 DPG / *** Joint with CB
Mandates can be found [here](#).



E responsabilità
italiane

[Link here](#)

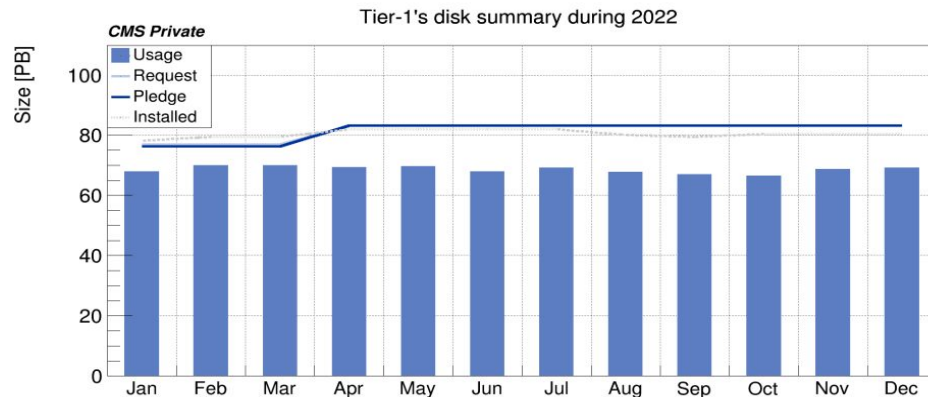
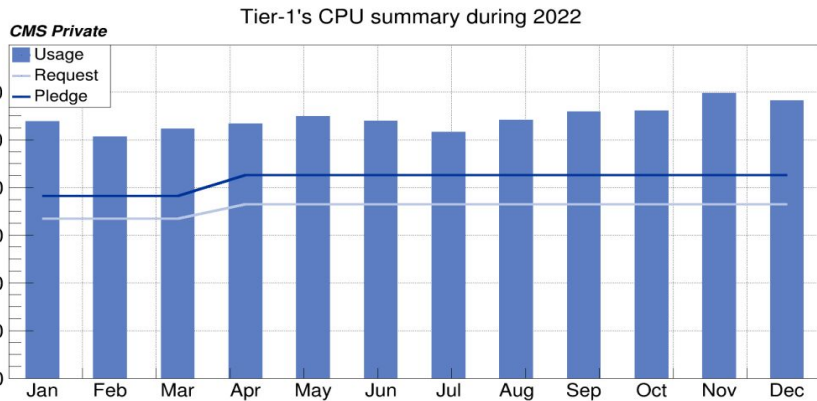
Backup



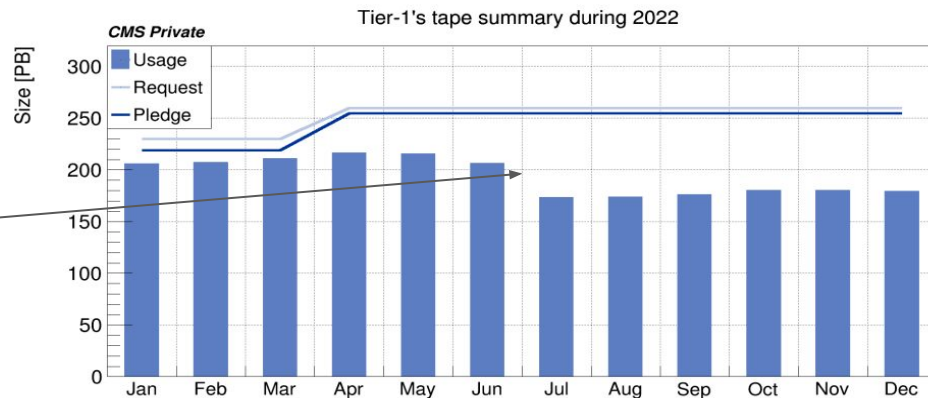
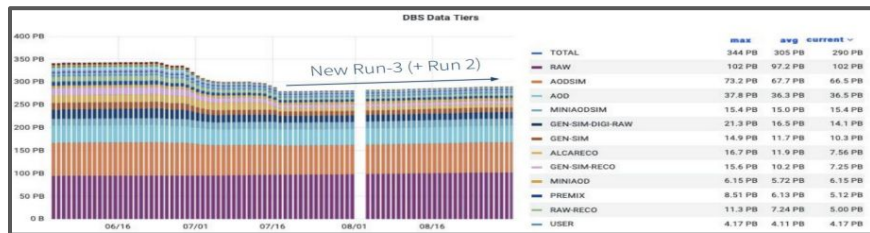
Prestazioni T1 CMS



Walltime Power [kHS06]



Tape Deletion Campaign (100PB)



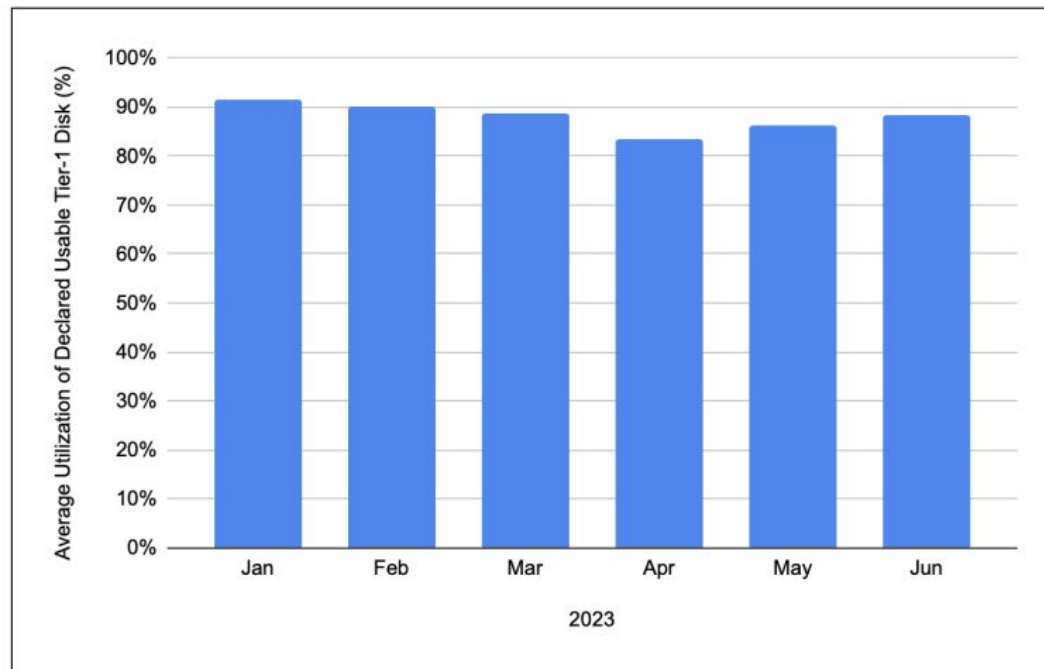
Uso disco @ Tier1s



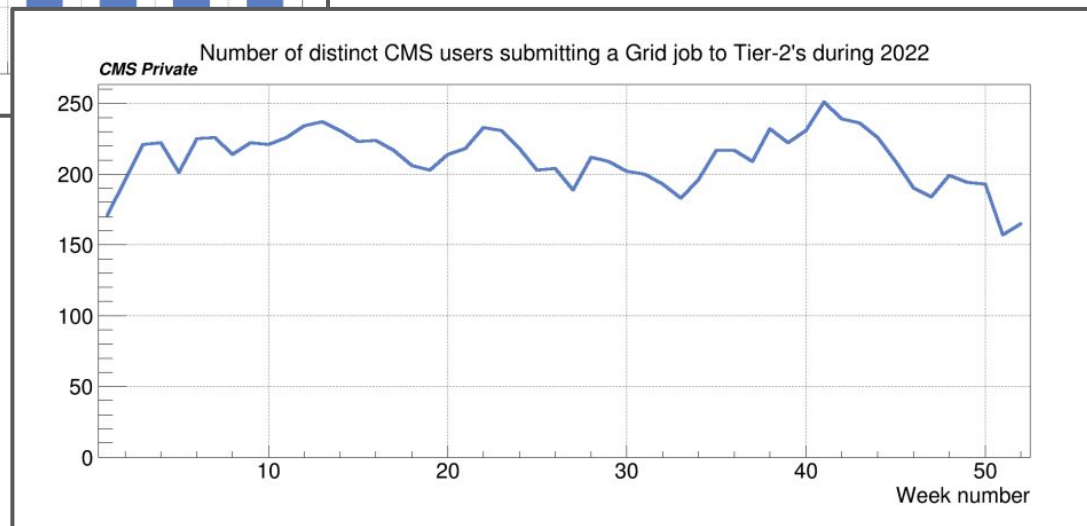
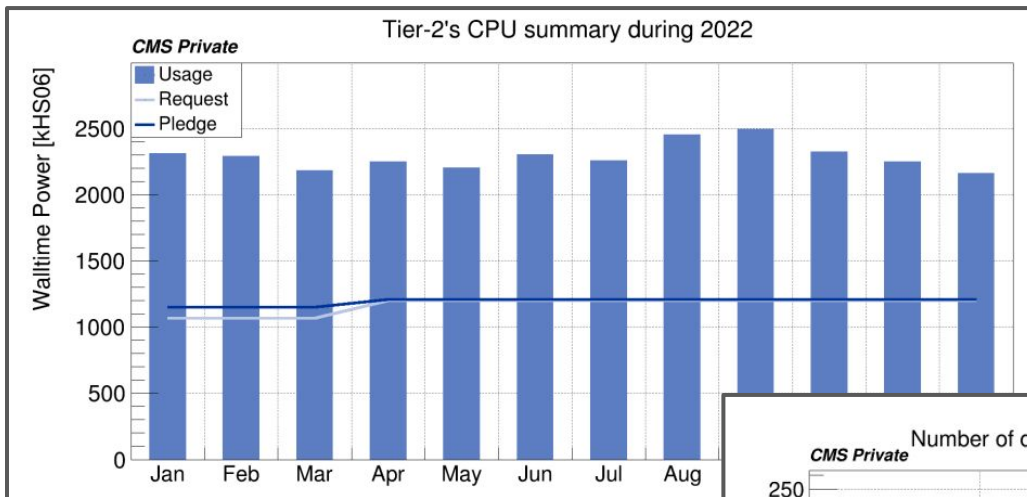
Rapporto “vero”...

Utilizzato VS realmente
disponibile

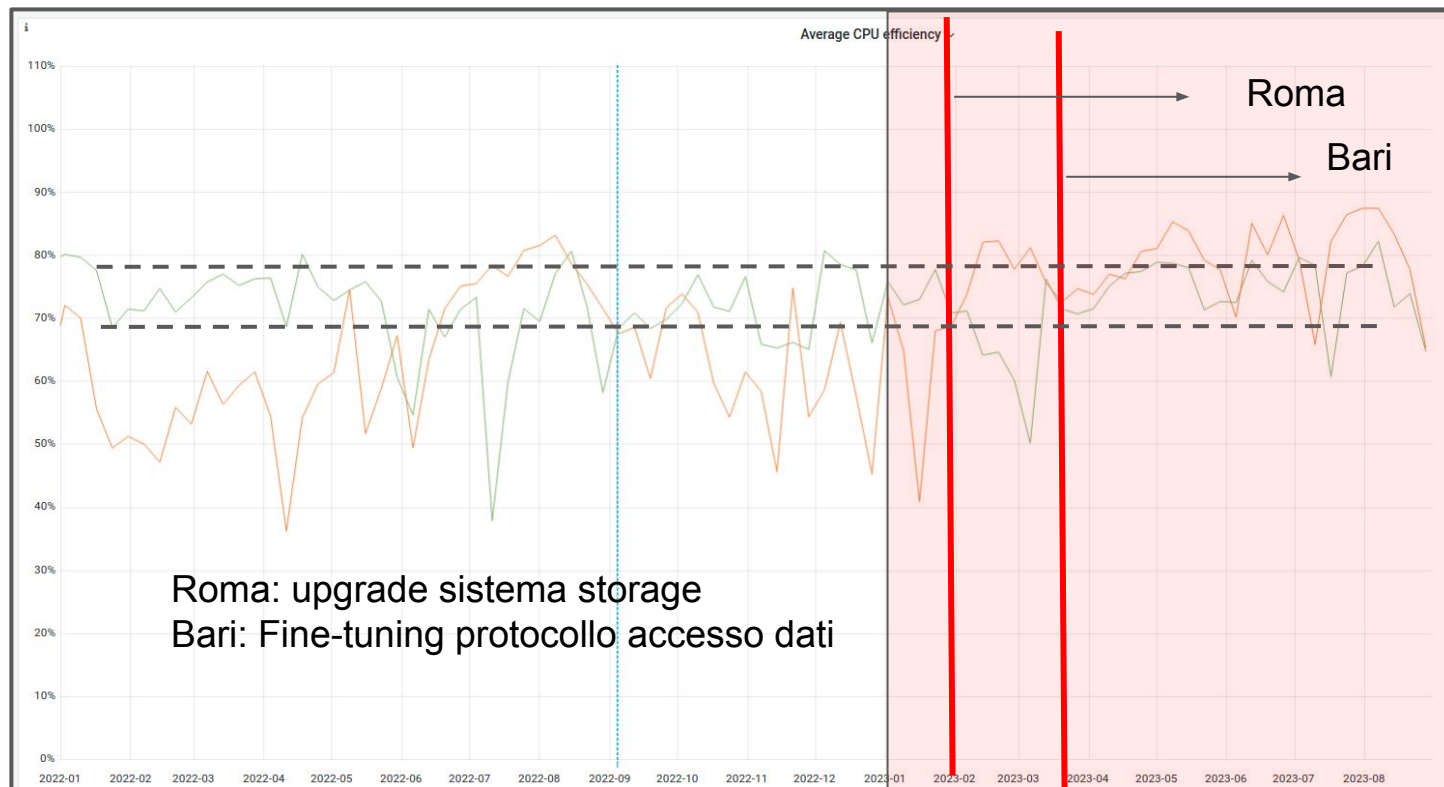
- E non rispetto al
valore pledged



Prestazioni T2 CMS



Evoluzione efficienza CPU Bari e Roma



Nota: Frazione CMS in crescita

PhD Scientists per Funding Agency Based on the Annually Revised Annex 13 of the M&O MoU

The List of Names is available at

https://cms-docdb.cern.ch/cgi-bin/DocDB/RetrieveFile?docid=3038&filename=CMS_PhD_List_2022.pdf&version=16

(Count closed on 13 September, 2021)

Institute FA	PhD #	PhD %
Austria	14	0.97%
Belgium-FNRS	22	1.52%
Belgium-FWO	14	0.97%
Brazil-FAPESP	7	0.48%
Brazil-RENAFAE	21	1.45%
Bulgaria	10	0.69%
CERN	80	5.54%
China	37	2.56%
Colombia	5	0.35%
Croatia	11	0.76%
Cyprus	8	0.55%
Ecuador	2	0.14%
Egypt	3	0.21%
Estonia	7	0.48%
Finland	14	0.97%
France-CEA	12	0.83%
France-IN2P3	40	2.77%
Germany-BMBF	57	3.95%
Germany-Helmholtz	42	2.91%
Greece	17	1.18%
Hungary	10	0.69%
India	30	2.08%
Iran	6	0.42%
Ireland	1	0.07%
Italy	197	13.64%

ANNEX 1

PhD Scientists per Funding Agency Based on the Annually Revised Annex 13 of the M&O MoU

The List of Names is available at

https://cms-docdb.cern.ch/cgi-bin/DocDB/RetrieveFile?docid=3038&filename=CMS_PhD_List_2023.pdf&version=16

(Count closed on 28 September 2022)

Institute FA	PhD #	PhD %
Austria	14	0.99%
Belgium-FNRS	21	1.48%
Belgium-FWO	13	0.92%
Brazil-FAPESP	7	0.49%
Brazil-RENAFAE	19	1.34%
Bulgaria	10	0.70%
CERN	68	4.79%
China	29	2.04%
Colombia	5	0.35%
Croatia	9	0.63%
Cyprus	8	0.56%
Ecuador	2	0.14%
Egypt	3	0.21%
Estonia	7	0.49%
Finland	15	1.06%
France-CEA	12	0.85%
France-IN2P3	38	2.68%
Germany-BMBF	63	4.44%
Germany-Helmholtz	45	3.17%
Greece	17	1.20%
Hungary	10	0.70%
India	29	2.04%
Iran	6	0.42%
Ireland	1	0.07%
Italy	206	14.51%

R&D GPU



Il programma di R&D di CMS per il calcolo eterogeneo (online e offline) è ben stabilito e prosegue secondo i piani

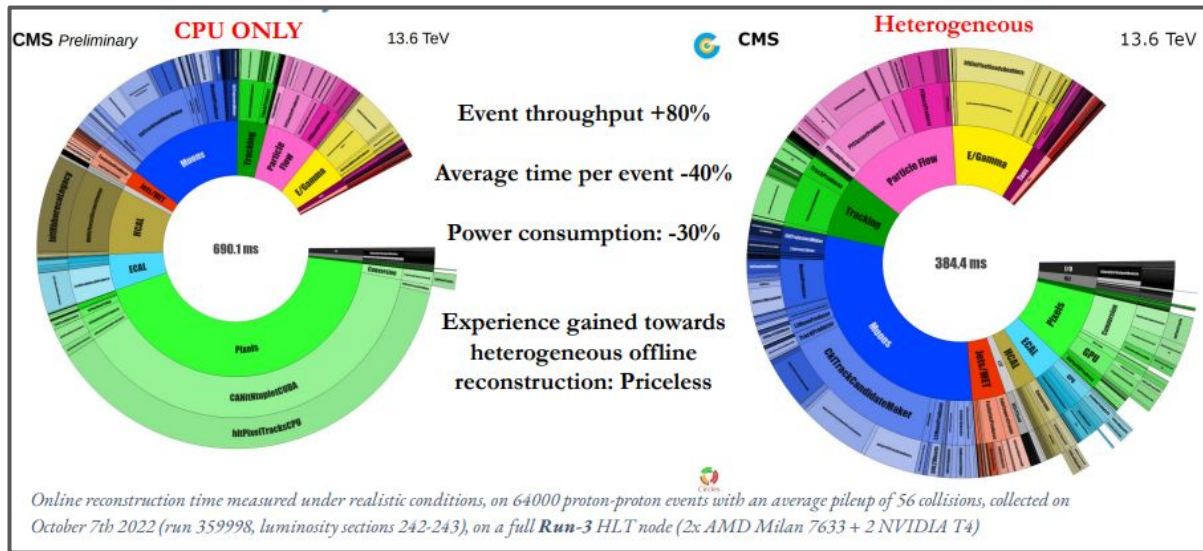
Oltre il 40% della ricostruzione **online** (HLT) di CMS è eterogenea

- Verrà interamente portato su Alpaka nel 2023
- Obiettivo dell'80% per la Fase 2



● Il 10% della ricostruzione **offline** del CMS sarà eterogeneo entro il 2023

- L'infrastruttura informatica esegue lavori di produzione eterogenei



VEGA- Estensione T1_CNAF...

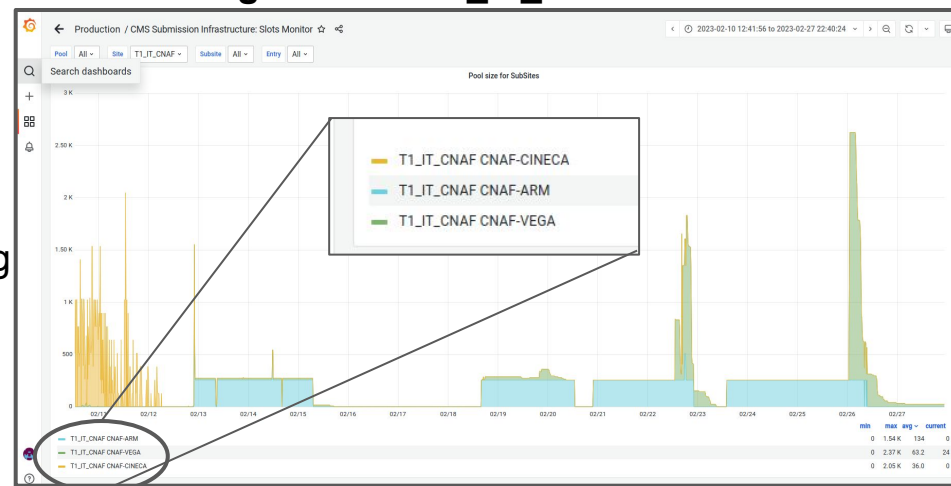


L'integrazione è stata fatta seguendo lo stesso approccio sviluppato e applicato all'estensione dei Tier1 su Marconi100



- Vega as **Subsite of T1_IT_CNAF**
 - Note “any” site could work
- **Manual pilots**
 - Pressure was generated in two different ways: via script on login node and also via [Virtual Kubelet](#) (the latter is something related to interTwin R&D)
- No local storage used at all
 - Inputs from AAA output to CNAF
- **CVMFS/Squids/Singularity (Apptainer) provided by Vega**

Running cores at T1_IT_CNAF subsites:



Attività sull'analisi ad alto rate (passato presente e futuro)



Finita la prima fase coi risultati presentati a [CHEP23](#) (e paper in review).

Ora iniziamo la seconda fase con lo studio dell'impatto su
Rete e Storage.



Prototyping a ROOT-based distributed analysis workflow for HL-LHC: the CMS case

Tommaso Tedeschi^{a,b,*}, Vincenzo Eduardo Maculano^c, Daniele Spiga^a, Diego Ciangottini^a, Mirco Tracoli^a, Eric Alejandro Saavedra^c, Enrico Guiraud^{c,e}, Massimo Biasotto^d

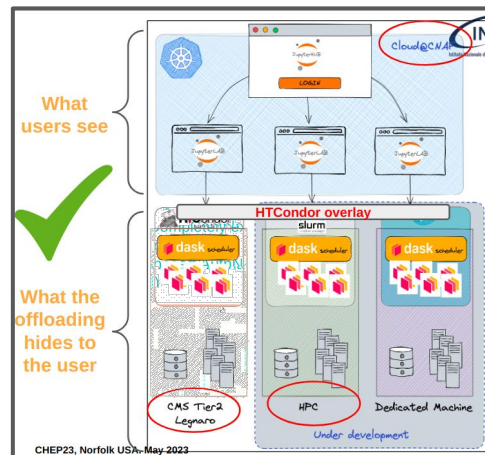
^aINFN Sezione di Perugia, Via A. Pascoli, 06123 Perugia, Italy

^bDepartment of Physics and Geology, University of Perugia, Via A. Pascoli, 06123 Perugia, Italy

^cEP-SFT, CERN, Meyrin, 1211 Geneva, Switzerland

^dINFN Laboratori Nazionali di Legnaro, Viale dell'Università 2, I-35020 Legnaro, Italy

^ePrinceton University, Princeton, NJ 08544, USA



<https://arxiv.org/abs/2307.12579>

Journal: Computer Physics
Communications

ppc64le commissioning @ Cineca M100



La disponibilità di risorse considerevoli di Power9 inizialmente presso CINECA è servita come spinta per adattare il codice WMAgent allo scenario “multiarch”

- Non solo supporta flussi di lavoro progettati per diversi architettura, ma supporta anche più di una architettura per workflow

L'impatto di risorse non x86_64 è ancora trascurabile, ma la funzionalità è essenziale i.e. per le attività emergenti con ARM (i.e. in particolare al CNAF)

Power architecture validation is completed

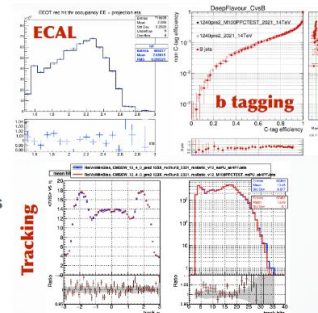
FullSim																
Release Name	Tracker	Ecal	HGcal	Hcal	CASTOR	DT	CSC	RPC	GEM	MTD	PPS	L1	Tracking	Electron	Photon	Muon
12_4_0_pre2_M100PPC	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

The Power CPU architecture can be now used for generation and processing of physics data.

- Both technical and physic validation have been completed
- We can now start using M100 HPC at CINECA for production activities

In addition:

- The underlying technical work (WM/SI mostly) has been beneficial to enable the support of multiple (any) architectures i.e. ARM [see later]
- Useful for the OLCF Summit (POWER9) allocation



CMS O&C Week, October 18-21, 2022

