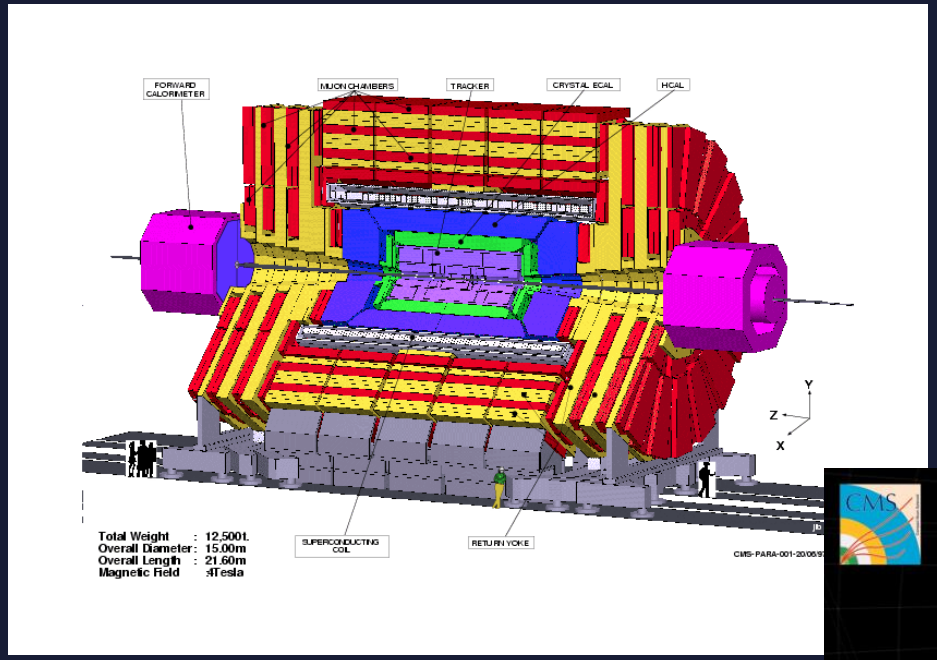
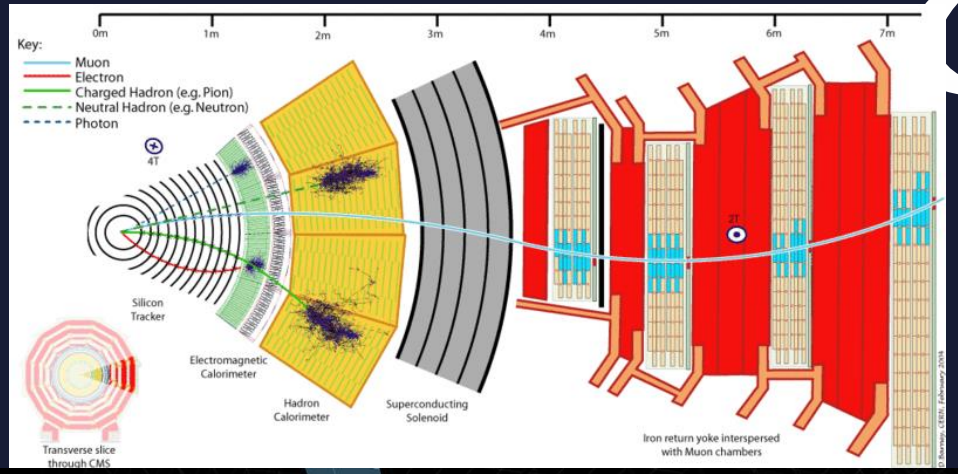
DataCloud – WP3
Risorse HTC e HPC

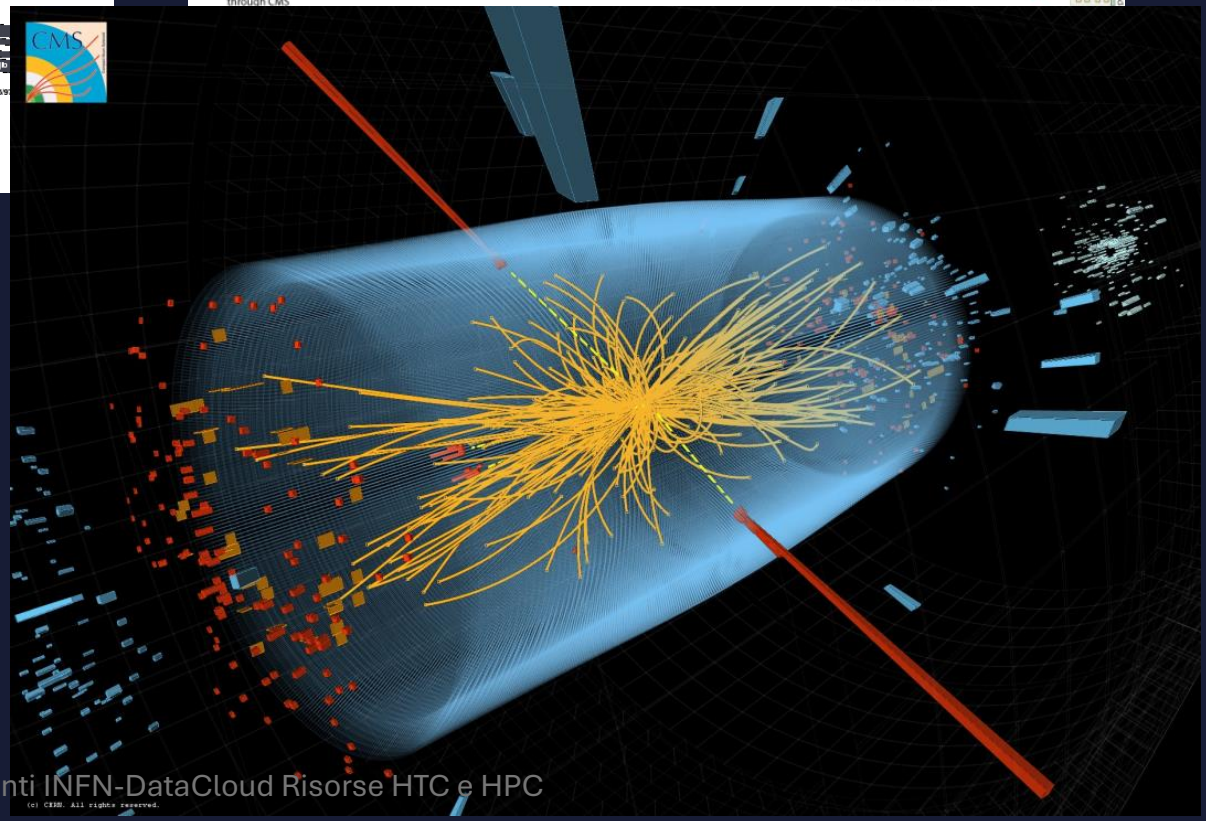
D.Cesini – INFN-CNAF

Outline

- HTC vs HPC
- La Grid
  - Perché federare le risorse
  - Evoluzione verso il Datalake
- Risorse pledged Al Tier-1 e ai Tier-2
  - CPU, DISK, TAPE
- HPC
- Le HPC Bubble


FORWARD CALORIMETER
MUON CHAMBERS
TRACKER
CRYSTAL ECAL
HCAL
SUPERCONDUCTING COIL
RETURN YOKE
Total Weight : 12,500t
Overall Diameter : 15.00m
Overall Length : 21.60m
Magnetic Field : 4Tesla
Y
Z
X


0m
1m
2m
3m
4m
5m
6m
7m
Key:
Muon
Electron
Charged Hadron (e.g. Pion)
Neutral Hadron (e.g. Neutron)
Photon
4T
2T
Silicon Tracker
Electromagnetic Calorimeter
Hadron Calorimeter
Superconducting Solenoid
Iron return yoke interspersed with Muon chambers
Transverse slice through CMS


CMS

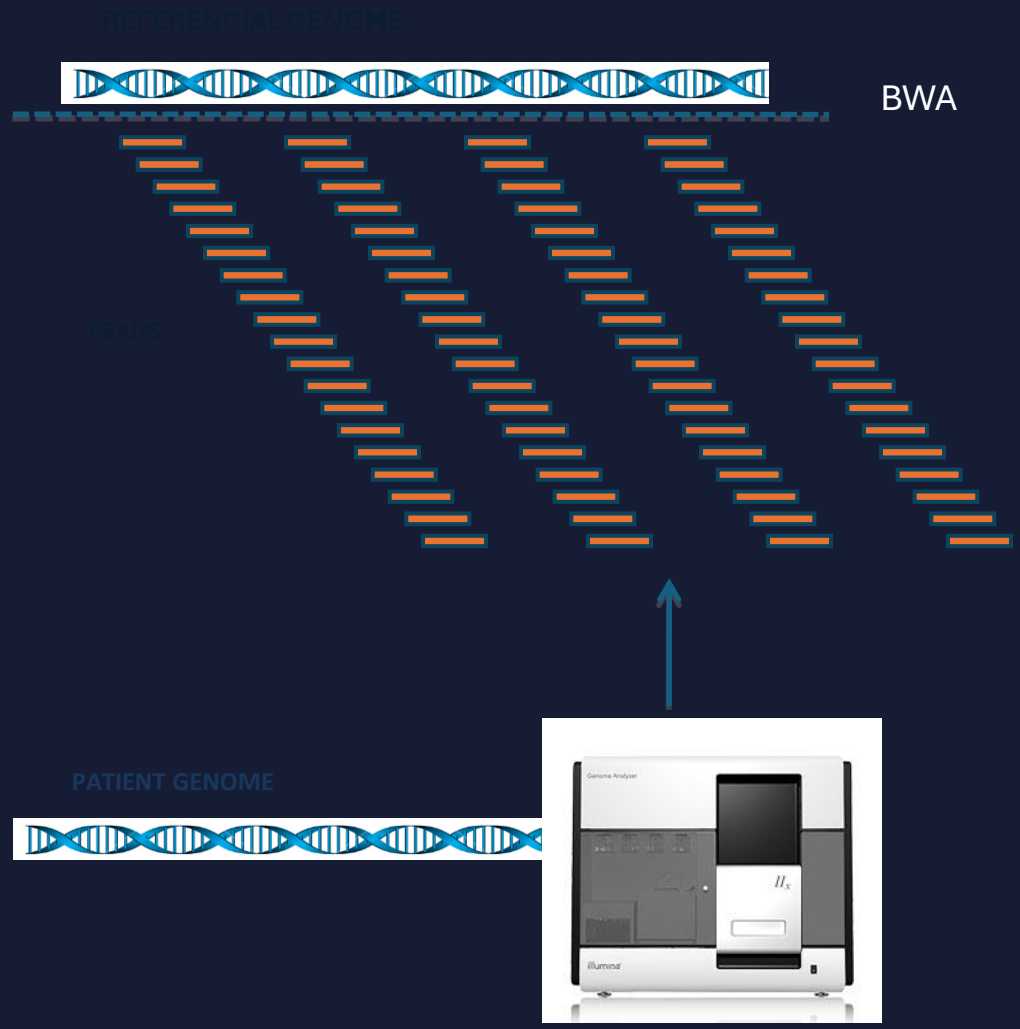
Massively Parallel Genome Sequencing

Used in the study of cancer Diseases

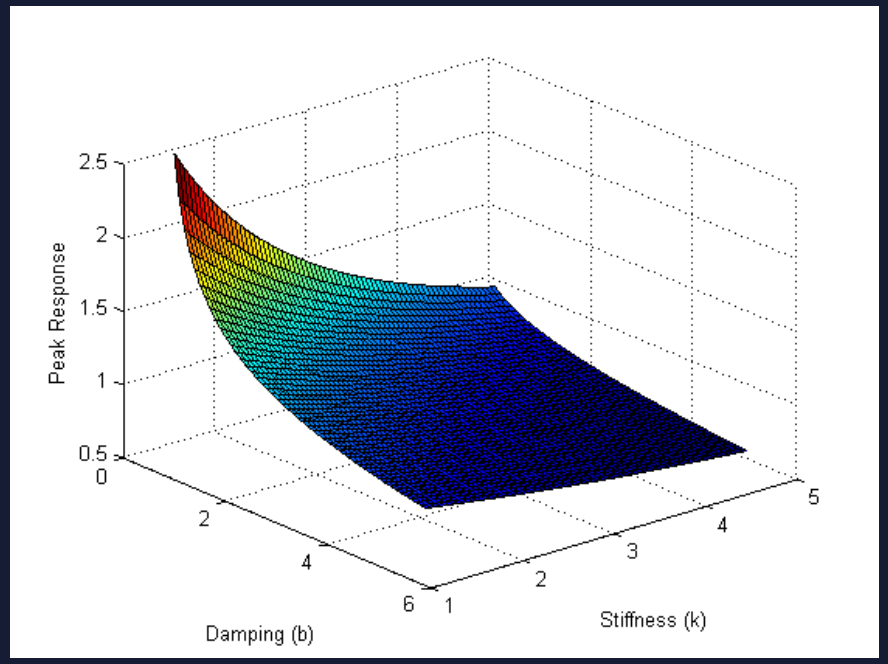
Allows massive amount of DNA or RNA fragments to be sequenced in a single experiment.

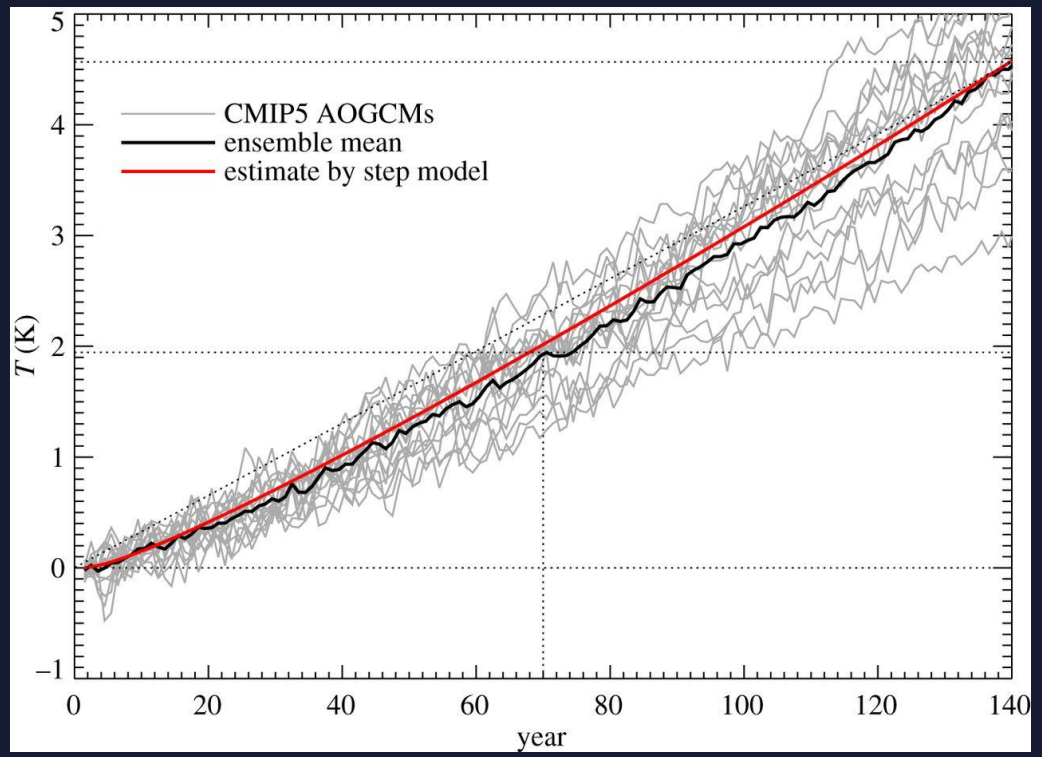
For the *massively parallel sequencing it is used* **BWA** tool (Burrows-Wheeler Aligner) for indexing and alignment

- Memory request~ 3,5 GB
- Total time ~ 50h using the group local resources



Parameter sweep and ensemble simulations





Grids and distributed systems

- What is a Grid?
- Grid types
- Anatomy of a Grid
- Accessing a Grid

© Grant Faint

What is a Grid? - Early definition

Ian Foster

I.Foster, C.Kesselman: The Grid: Blueprint for a New Computing Infrastructure", 1998

Carl Kesselman

"A computational Grid is a hardware and software infrastructure that provides dependable, consistent, pervasive and inexpensive access to high-end computational capabilities"

What is a computational Grid? the 3 points checklist

A Grid is a system that.....

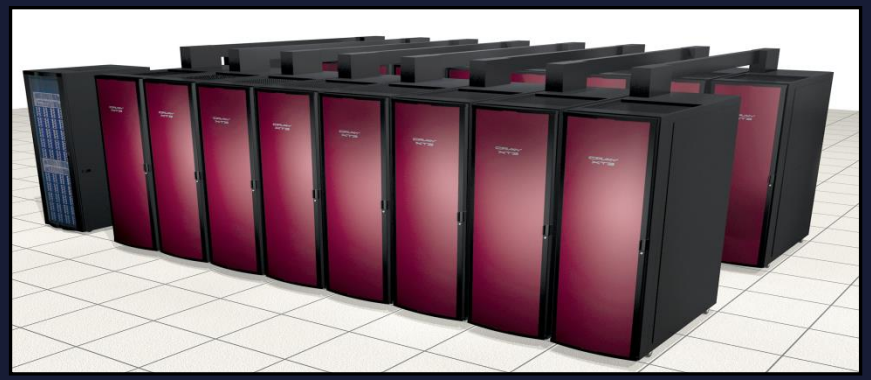
1) **Coordinates resources that are not subject to centralized control**
2) **Uses standard, open, general-purpose protocols and interfaces**
3) **Delivers nontrivial qualities of service** (Ian Foster, 2002)

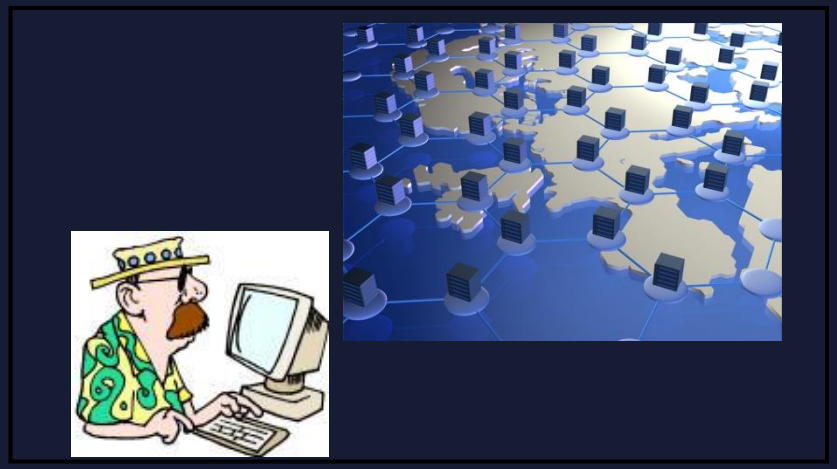
[1] Foster, I. and Kesselman, C. eds. The Grid: Blueprint for a New Computing Infrastructure, Morgan Kaufmann, 1999, 259-278
[2] Ian Foster, Carl Kesselman, and Steven Tuecke. 2001. The Anatomy of the Grid: Enabling Scalable Virtual Organizations. Int. J. High Perform. Comput. Appl. 15, 3 (August 2001), 200-222. DOI=10.1177/109434200101500302
[3] What is the Grid? A Three Point Checklist. I. Foster, GRIDToday, July 20, 2002.

Grid: No centralized control

The user in general has full ownership of a desktop workstation.

A Cluster is a shared resource – Only the administrator has full control of the system
The physical layer is still well defined.





I submit my jobs to “the GRID” and they get processed: somehow, somewhere, after some time.

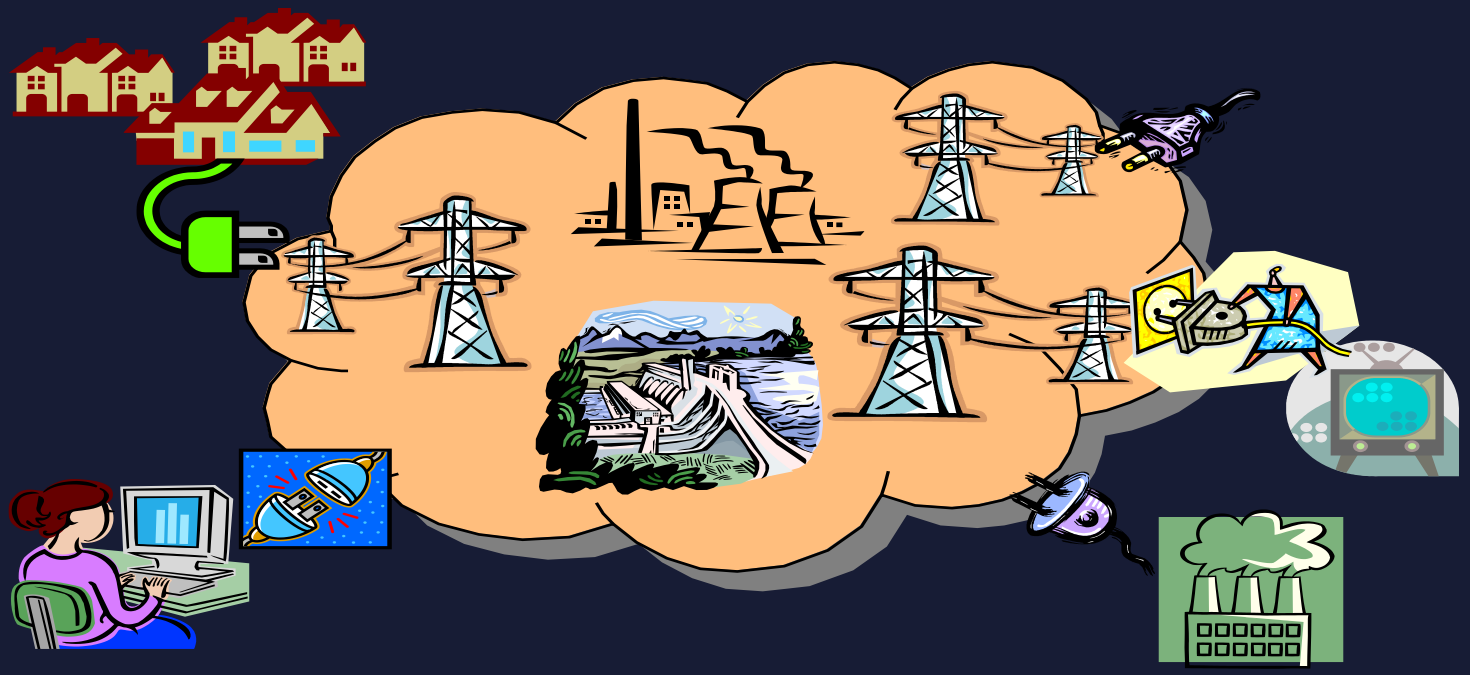
There is no GRID owner!

1st Law of the Grid

- **95% of the Grid is… agreement**

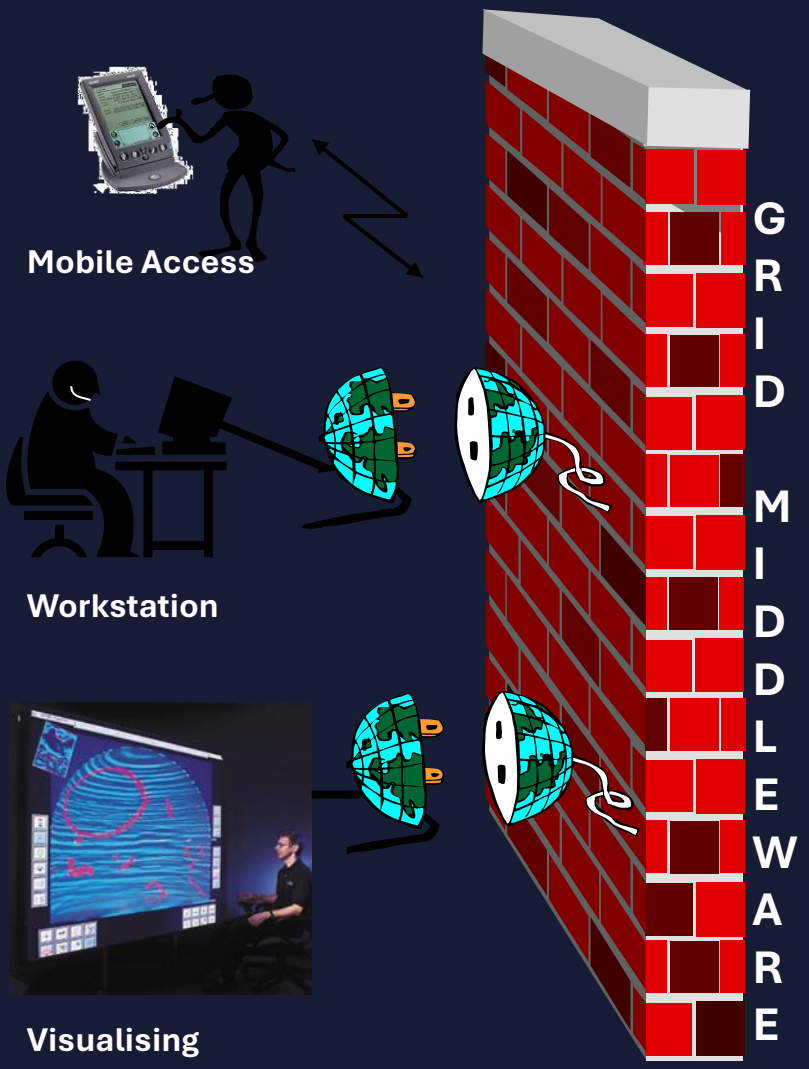
- Key terms
  - Coordination
  - No centralized control
  - Standards
  - Protocols
  - Interfaces
- Standards, protocols, interfaces,… aim at providing common abstractions of different implementations of similar services

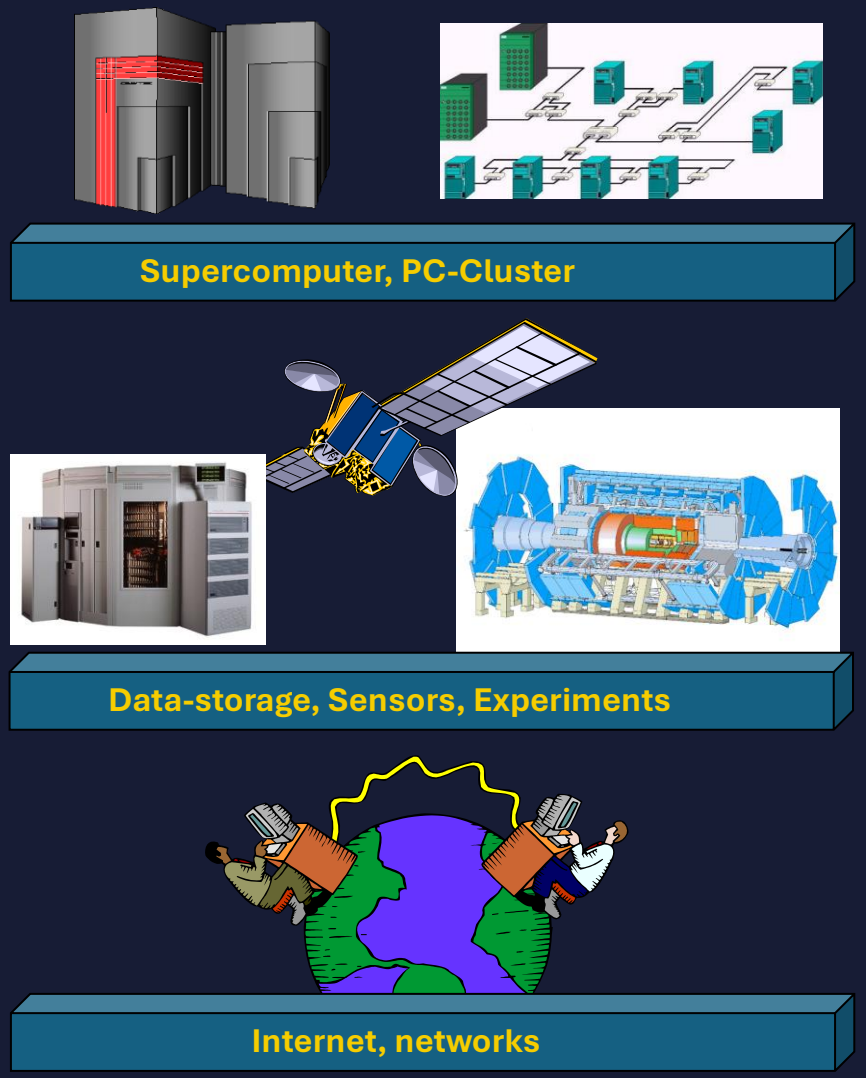
Power Grid Similarity



“We will probably see the spread of computer utilities, which, like present electric and telephone utilities, will service individual homes and offices across the country”
(Len Kleinrock, 1969)

The Grid Paradigm





Grid Types

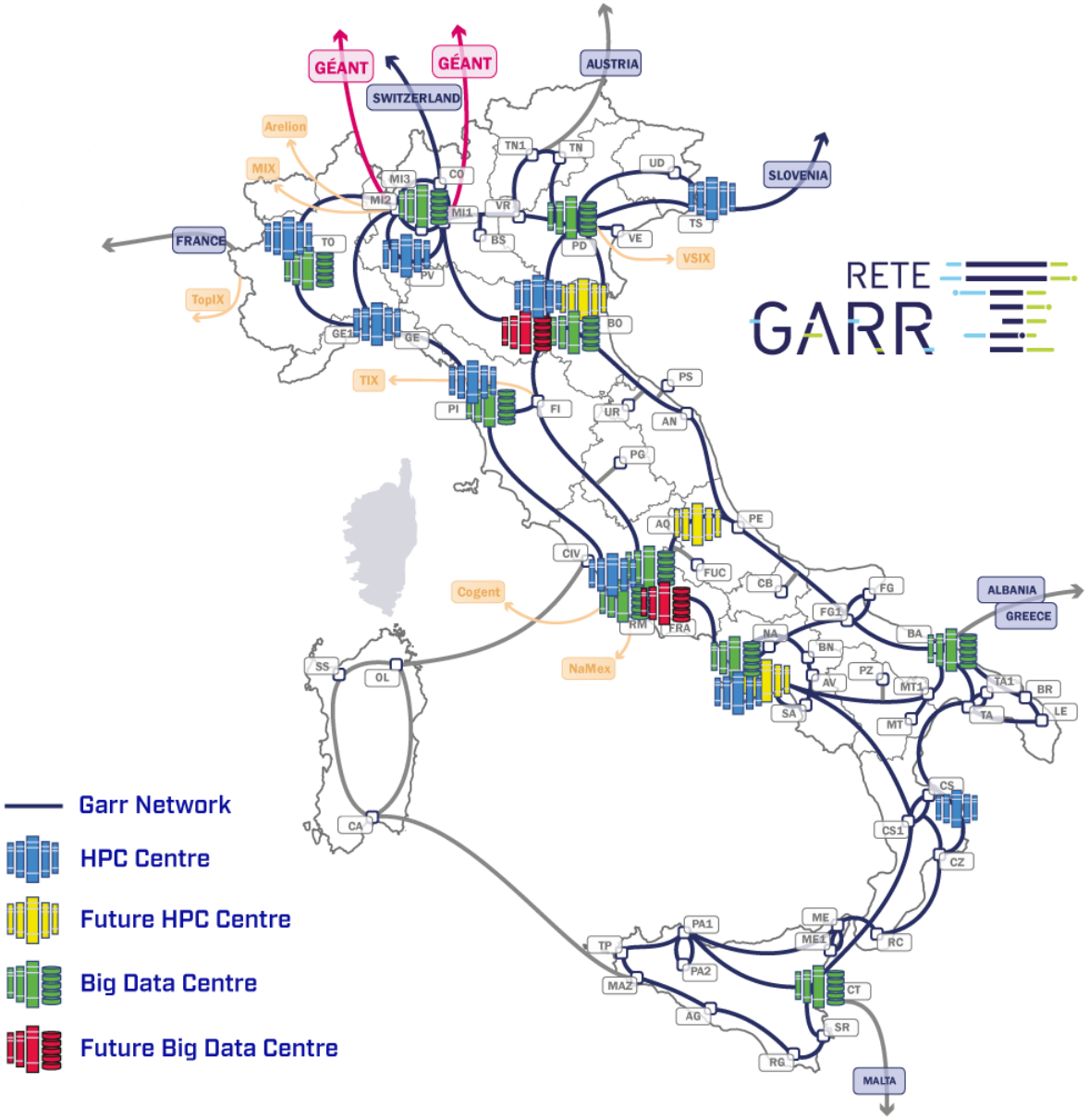
Supercomputer Based Service Grid

Cluster Based Service Grid

Volunteer Desktop Grid

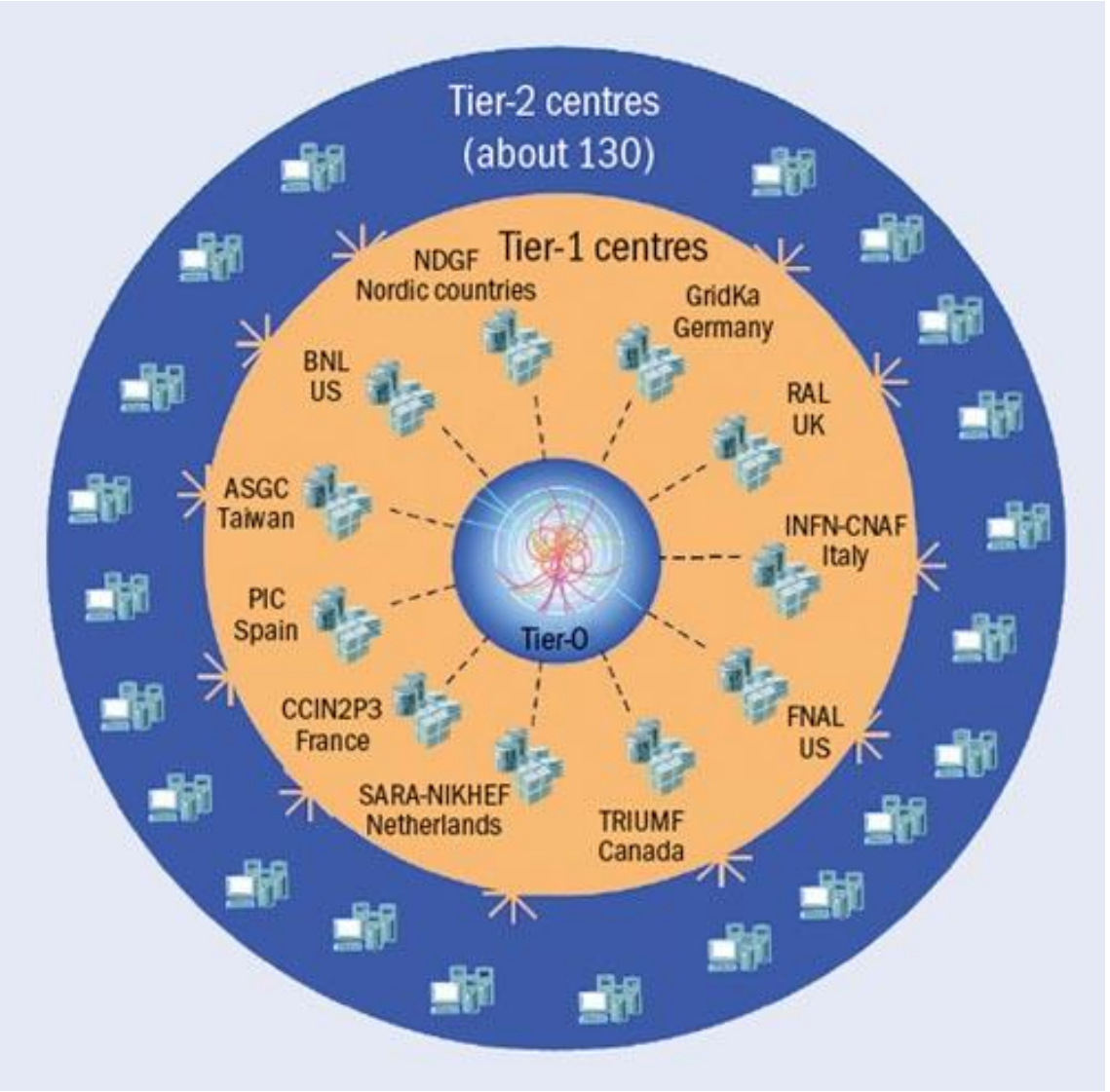
ICSC SPOKE 0
Infrastruttura
Cloud di supercalcolo

ICSC
Centro Nazionale di Ricerca in HPC,
Big Data and Quantum Computing

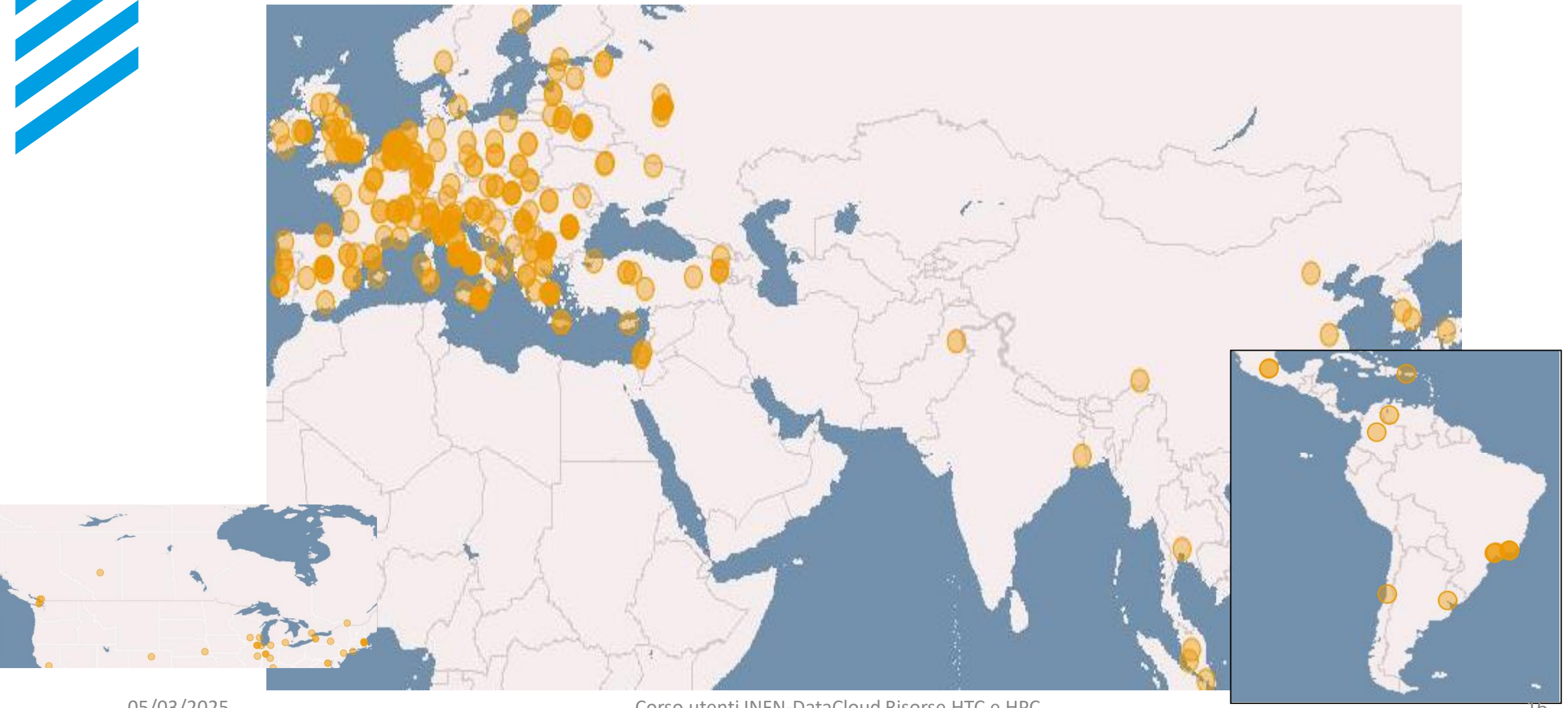


WLCG

- **Worldwide LHC computing Grid**
- Service GRID for the LHC high energy physics experiments
- Tiered structure
- Part of the European grid Infrastructure (EGI)
  - O(1M) logical CPUs
  - O(1) EB disk
  - O(1) EB tape



The European Grid Infrastructure



The European Grid Infrastructure

A typical Grid site

Batch System (PBS,LSF…)

Computing Element (CE)

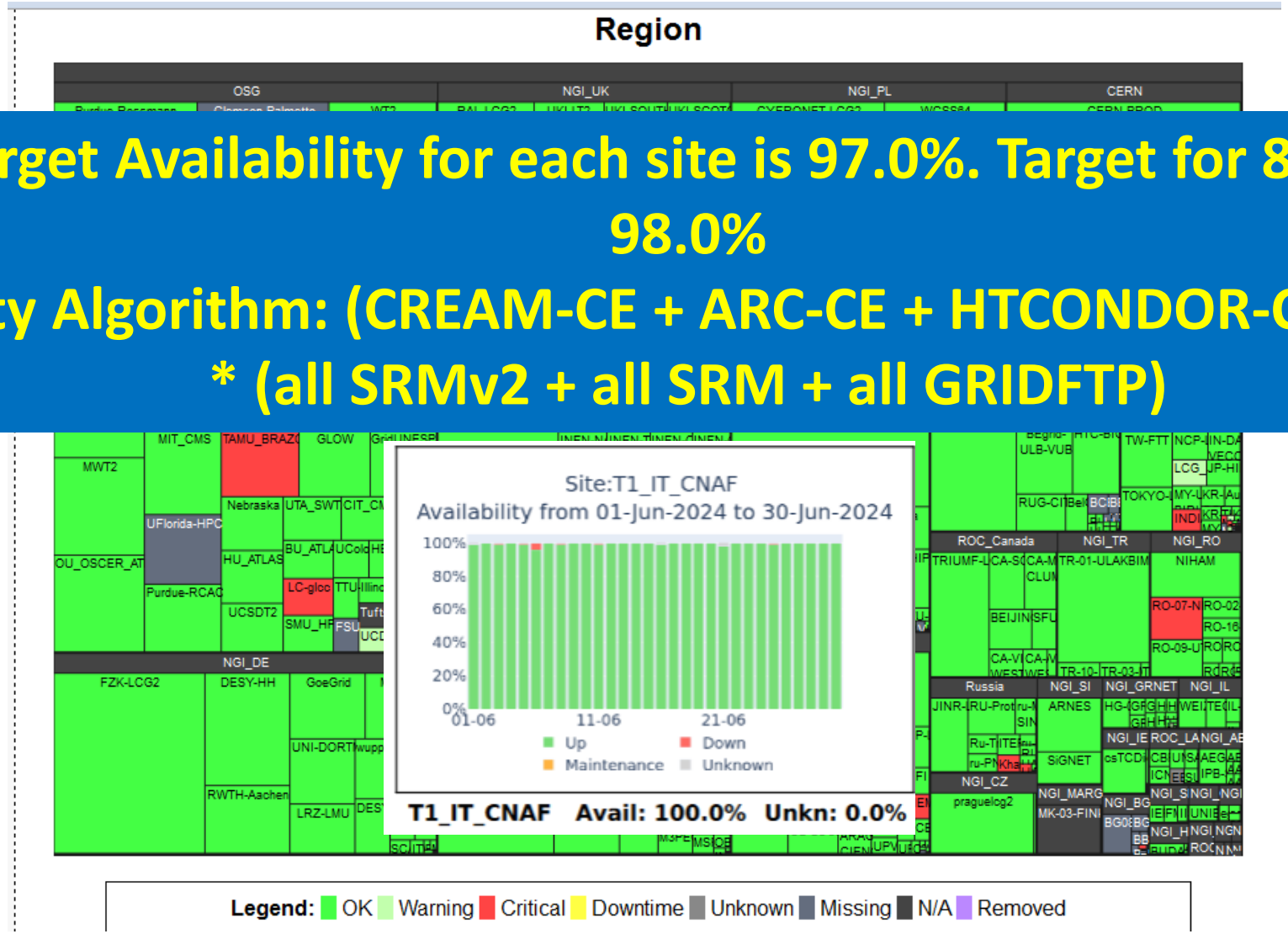
Storage Element (SE)

INFORMATION SYSTEM (es. BDII)

Operations Tools: Monitoring

WLCG Target Availability for each site is 97.0%. Target for 8 best sites is 98.0%

Availability Algorithm: (CREAM-CE + ARC-CE + HTCONDOR-CE + GLOBUS) * (all SRMv2 + all SRM + all GRIDFTP)



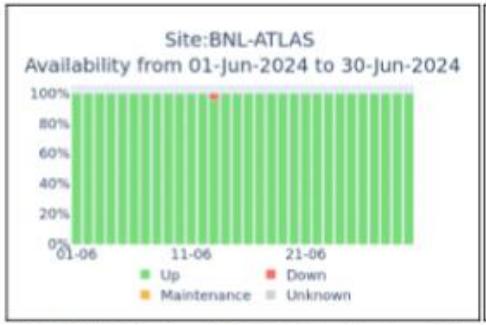
C

Availability of WLCG Tier-0 + Tier-1 Sites

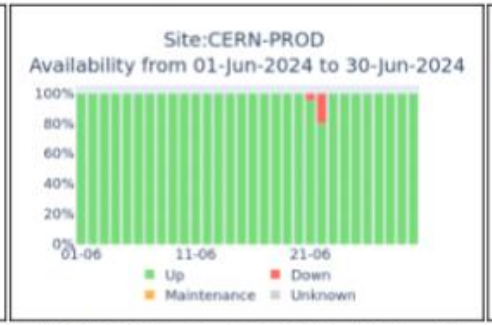
ATLAS **June 2024**

Target Availability for each site is 97.0%. Target for 8 best sites is 98.0%

Availability Algorithm: (CREAM-CE + ARC-CE + HTCONDOR-CE + GLOBUS) * (all SRMv2 + all SRM + all GRIDFTP)



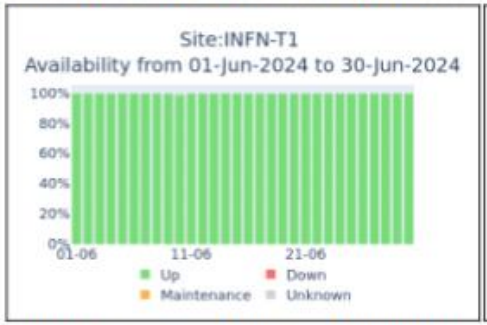
BNL-ATLAS Avail: 100.0% Unkn: 0%



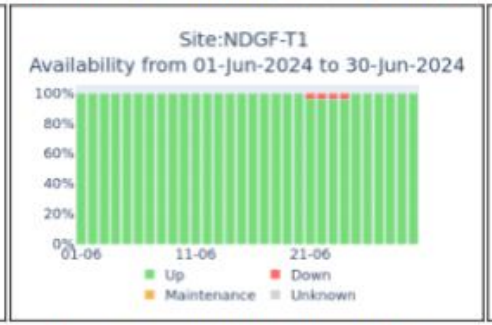
CERN-PROD Avail: 99.0% Unkn: 0%

FZK-LCG2 Avail: 100.0% Unkn: 0%

IN2P3-CC Avail: 100.0% Unkn: 0%



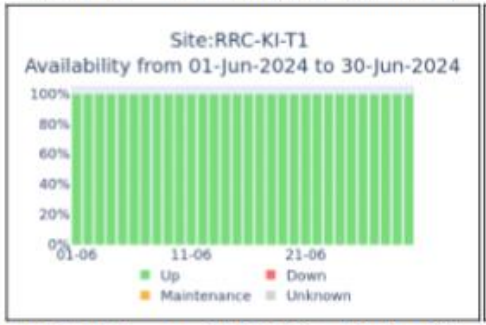
INFN-T1 Avail: 100.0% Unkn: 0.0%



NDGF-T1 Avail: 99.0% Unkn: 0%

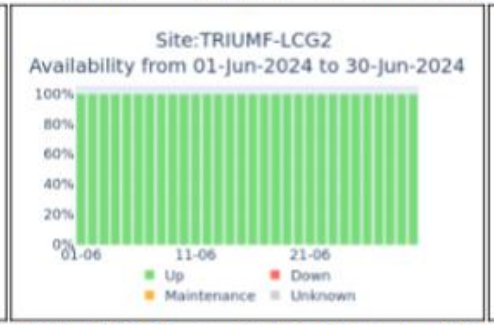
NIKHEF Avail: 93.0% Unkn: 0%

RAL-LCG2 Avail: 100.0% Unkn: 0%

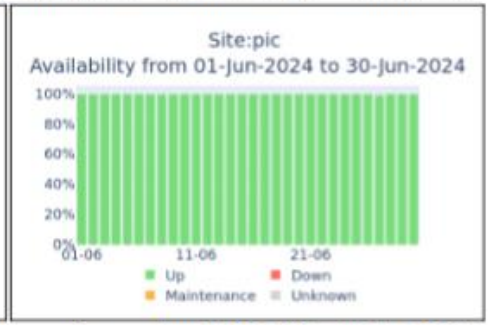


RRC-KI-T1 Avail: 100.0% Unkn: 0%

SARA-MATRIX Avail: 95.0% Unkn: 0.0%



TRIUMF-LCG2 Avail: 100.0% Unkn: 0%



pic Avail: 100.0% Unkn: 0.0%

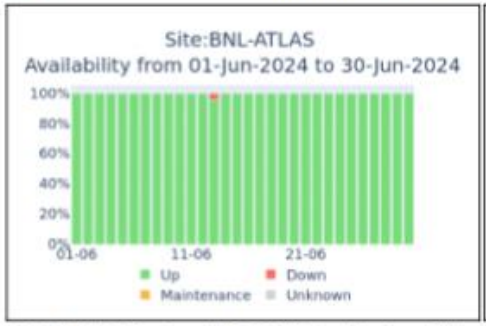
C

Availability of WLCG Tier-0 + Tier-1 Sites

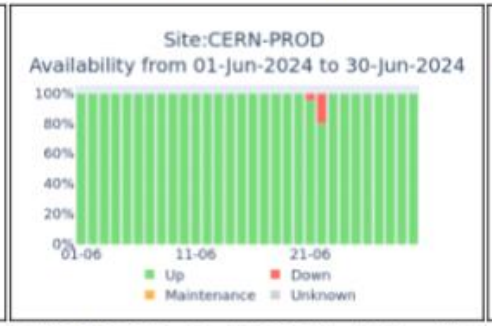
ATLAS **June 2024**

Target Availability for each site is 97.0%. Target for 8 best sites is 98.0%

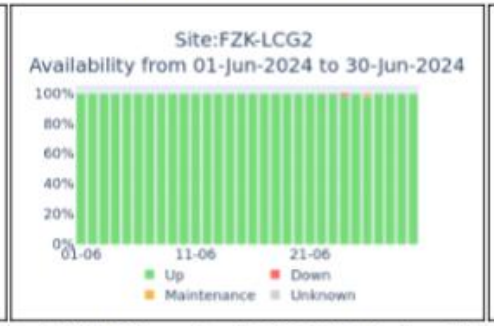
Availability Algorithm: (CREAM-CE + ARC-CE + HTCONDOR-CE + GLOBUS) * (all SRMv2 + all SRM + all GRIDFTP)



BNL-ATLAS Avail: 100.0% Unkn: 0%

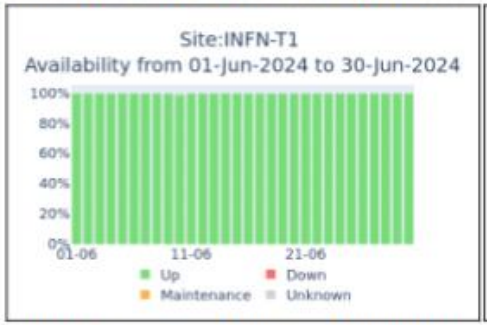


CERN-PROD Avail: 99.0% Unkn: 0%

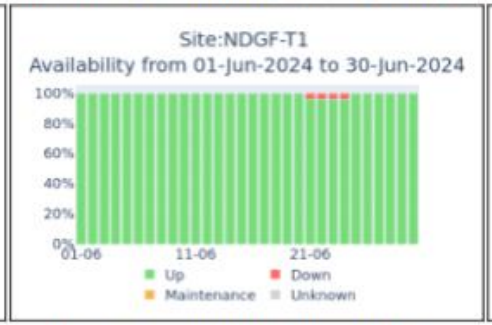


FZK-LCG2 Avail: 100.0% Unkn: 0%

IN2P3-CC Avail: 100.0% Unkn: 0%



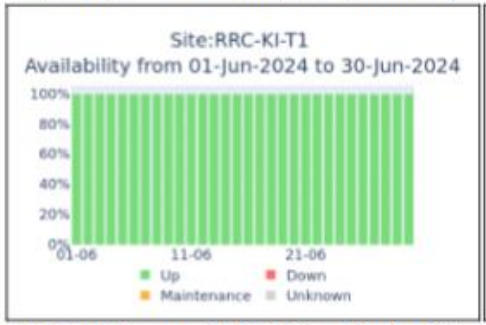
INFN-T1 Avail: 100.0% Unkn: 0.0%



NDGF-T1 Avail: 99.0% Unkn: 0%

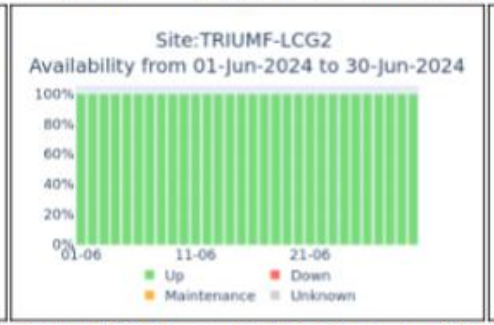
NIKHEF Avail: 93.0% Unkn: 0%

RAL-LCG2 Avail: 100.0% Unkn: 0%

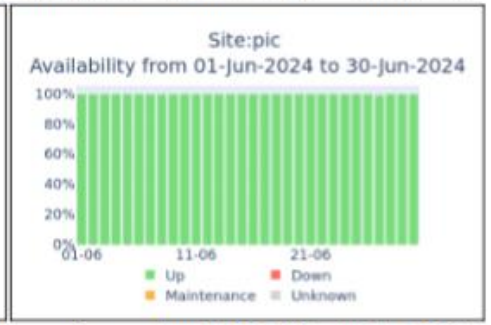


RRC-KI-T1 Avail: 100.0% Unkn: 0%

SARA-MATRIX Avail: 95.0% Unkn: 0.0%



TRIUMF-LCG2 Avail: 100.0% Unkn: 0%



pic Avail: 100.0% Unkn: 0.0%

Operations Tools: Monitoring

Tier-2 Availability and Reliability Report
ATLAS **June 2024**

Federation Summary - Sorted by Availability

Availability Algorithm: (CREAM-CE + ARC-CE + HTCONDOR-CE + GLOBUS) * (all SRMv2 + all SRM + all GRIDFTP)

Color coding: N/A <30% <60% <90% >=90%

Federation	Availability	Reliability
PL-POLISH-WLCG	100%	100%
IL-HEPTier-2	100%	100%
US-NET2	100%	100%
CN-IHEP	100%	100%
HK-ATLAS-T2	100%	100%
PT-LIP-LCG-Tier2	100%	100%
CA-EAST-T2	100%	100%
FR-IN2P3-CPPM	100%	100%
FR-IN2P3-LAPP	99%	99%
JP-Tokyo-ATLAS-T2	99%	100%
CH-ATLAS	99%	99%
SI-SiGNET	99%	99%
US-MWT2	99%	99%
US-AGLT2	99%	99%
FR-IN2P3-LPC	98%	98%
RU-RDIG	98%	98%
FR-GRIF	97%	97%
IT-INFN-T2	97%	97%

Federation	Availability	Reliability
UK-London-Tier2	97%	97%
ES-ATLAS-T2	97%	98%
T2-LATINAMERICA	96%	96%
DE-FREIBURGWUPPERTAL	95%	100%
CA-WEST-T2	94%	98%
DE-DESY-ATLAS-T2	94%	98%
RO-LCG	91%	91%
UK-SouthGrid	89%	96%
DE-DESY-GOE-ATLAS-T2	88%	91%
US-SWT2	88%	88%
SK-Tier2-Federation	86%	86%
DE-MCAT	85%	85%
CZ-Prague-T2	81%	96%
UK-NorthGrid	75%	75%
TR-Tier2-federation	70%	73%
UK-ScotGrid	22%	28%
CH-CHIPP-CSCS	1%	3%
SE-SNIC-T2	0%	0%

Perché infrastrutture federate

- Non è una questione di ottenere una infrastruttura più grande da una serie di centri di calcolo
- Valore aggiunto nella necessità di avvicinare le comunità
- Supporto a comunità internazionali e distribuite per natura
  - Creazione di Virtual Organization
- Failover e disaster recovery

2nd Law of the Grid

Anything that can go wrong, will

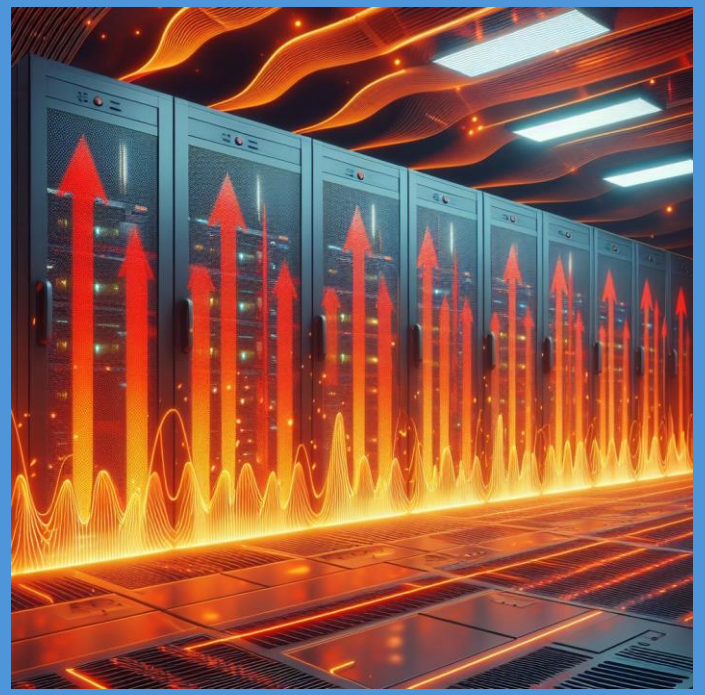
- The Grid is a very complex environment, errors will happen
- Some errors are preventable, some are manageable by the infrastructure, some can only be managed by the user

“A distributed system is one in which the failure of a computer you didn‘t even know existed can render your own computer unusable.”
Leslie Lamport

Expect the unexpected

When the Air-conditioning / power fails (again & again & again);

©Jamie Shiers 2008 J. Phys.: Conf. Ser. 119 052030 – Lessons Learnt from WLCG Service Deployment



Generated with AI · May 2024

Expect the unexpected

When a service engineer puts a Coke into a machine to 'warm it up';

Generated with AI · May 2024

©Jamie Shiers 2008 J. Phys.: Conf. Ser. 119 052030 – Lessons Learnt from WLCG Service Deployment

Expect the unexpected

When a fishing trawler cuts a trans-Atlantic network cable;

Generated with AI · May 2024

©Jamie Shiers 2008 J. Phys.: Conf. Ser. 119 052030 – Lessons Learnt from WLCG Service Deployment

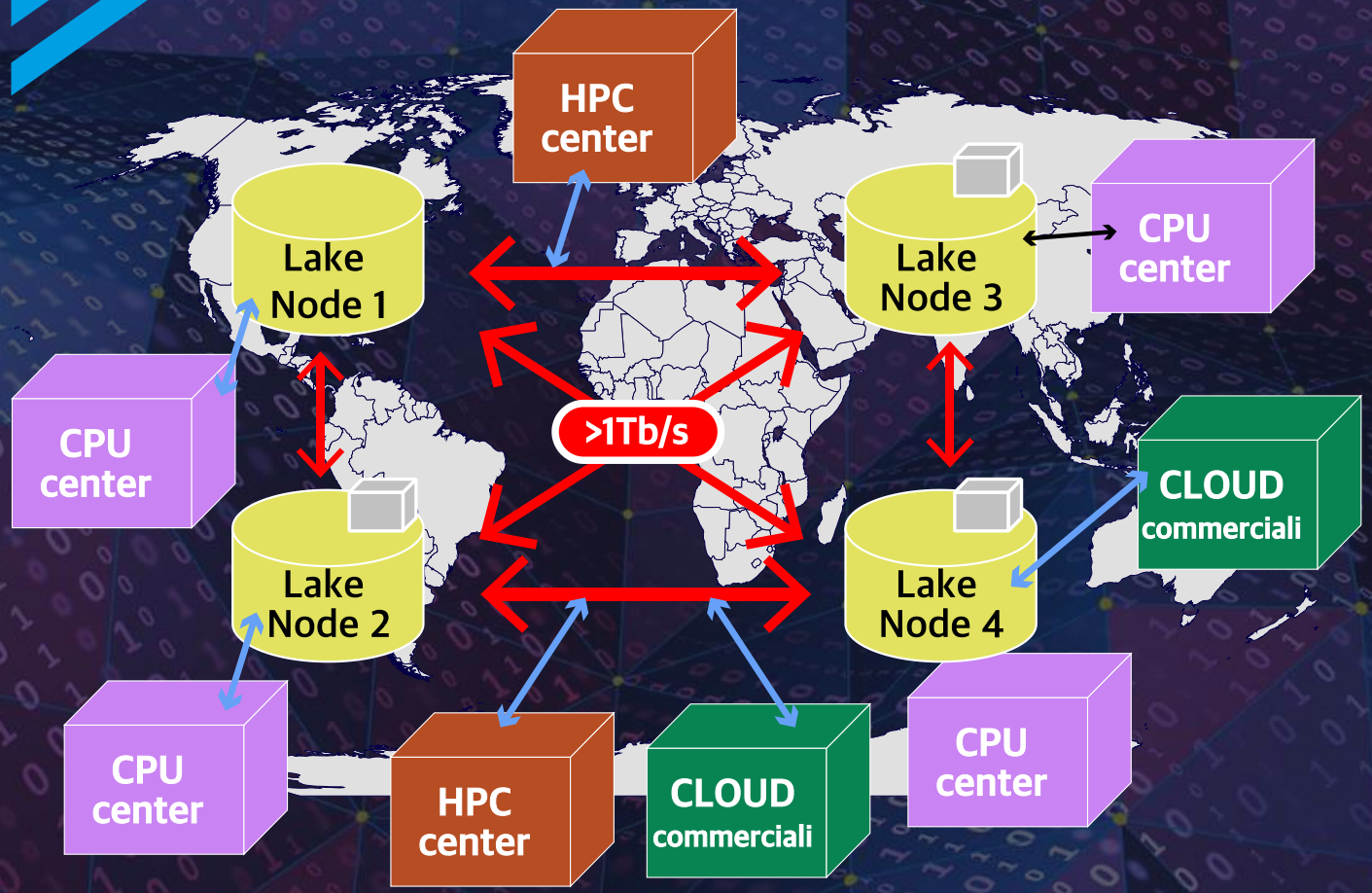
Expect the unexpected

> *When a Tsunami does the equivalent in Asia Pacific;*

Generated with AI · May 2024

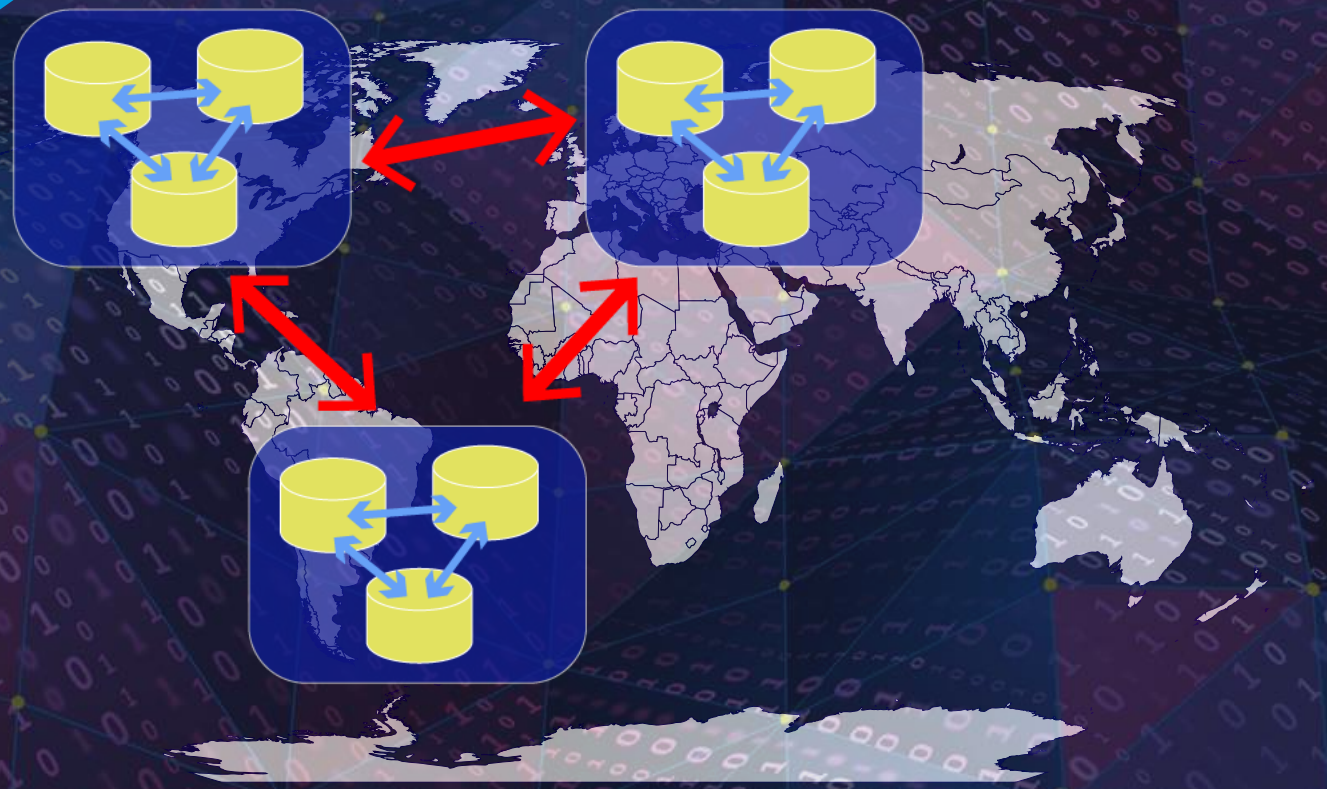
©Jamie Shiers 2008 J. Phys.: Conf. Ser. 119 052030 – Lessons Learnt from WLCG Service Deployment

Il **datalake** scientifico e la sua evoluzione



- Struttura logicamente singola
  - Interconnessa tramite reti ad alta banda multi-Tbps “inter-lake”
- Disaccoppiamento completo degli aspetti di storage e calcolo
- Elevata disponibilità e sicurezza del dato in centri dedicati
  - Minimizzazione del numero di copie
- CPU/GPU: utilizzate dovunque si trovino agganciandole al «Lake»
- Il datalake sfruttabile da altre comunità nazionali ed internazionali

Il **datalake** scientifico e la sua evoluzione



- Struttura logicamente singola
  - Interconnessa tramite reti ad alta banda “inter-lake”
- Disaccoppiamento completo degli aspetti di storage e calcolo
- Elevata disponibilità e sicurezza del dato in centri dedicati
  - Minimizzazione delle copie
- CPU/GPU: utilizzate dovunque si trovino agganciandole al «Lake»
- Il datalake sfruttabile da altre comunità nazionali ed internazionali

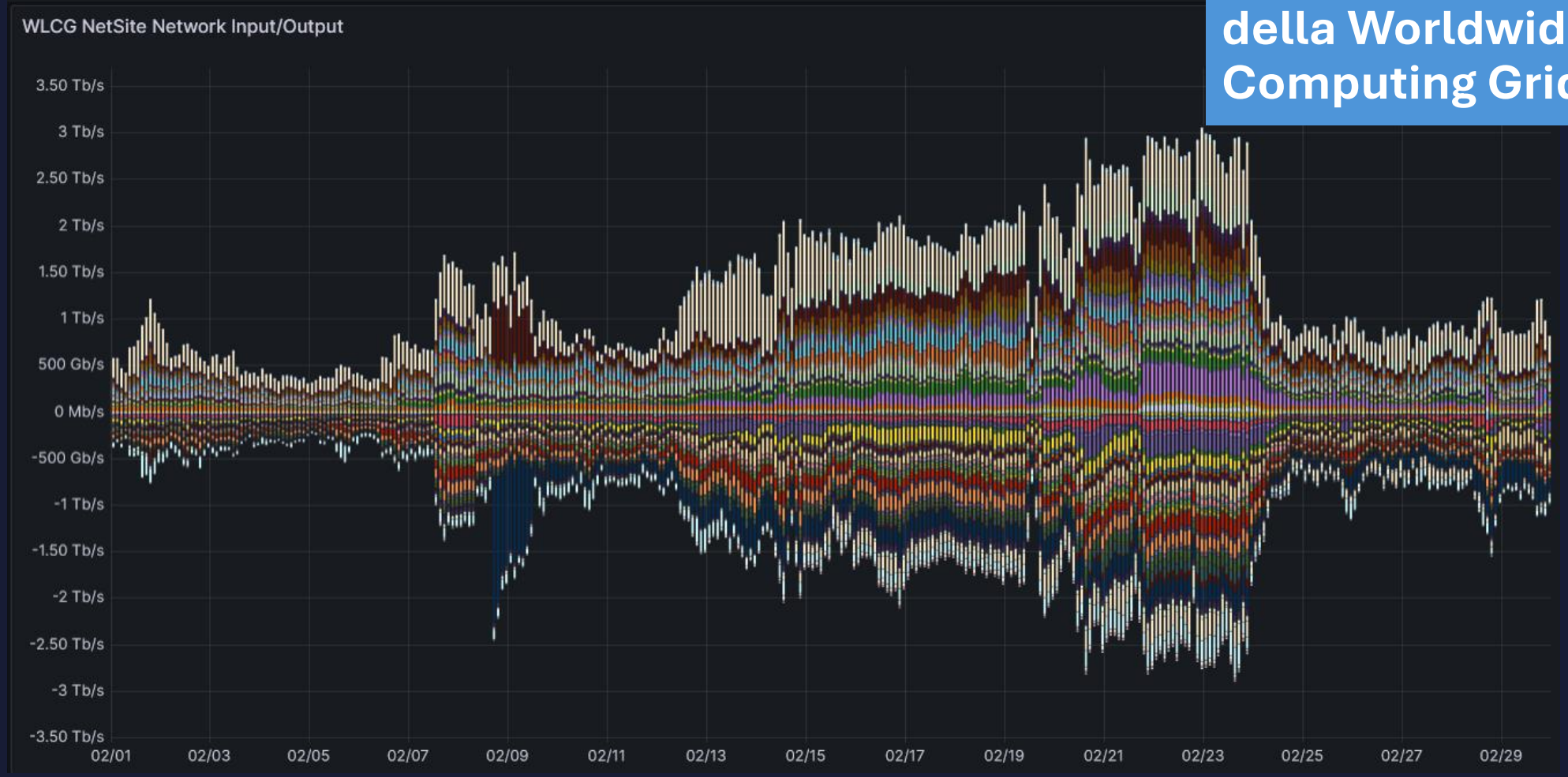
Generated with AI · May 2024

Come dovremmo fare il **Datalake <u>secondo</u> l'Intelligenza Artificiale**

Draw ***«a worldwide datalake with high speed network connections»***

Lo sviluppo dei sistemi di gestione dei dati in un mondo data-intensive

La strada verso la realizzazione del Data Lake
II Data Challenge 2024 della Worldwide LHC Computing Grid



La gestione dei dati in un mondo sempre più **data-intensive**

Data challenge al CNAF effettuato attraverso il nuovo link ottico diretto con il CERN

- Estensione diretta delle rispettive reti su spettro condiviso multi-dominio
- Alta banda: attualmente 400 Gbps, fino a 1.6 Tbps; Latenza: 9.5s

289 Gbps

Resources@Tier1

Le risorse PLEDGED

PLEDGE @T1 2024-2025

Experiment	01/04/2024			01/01/2025		
	CPU	DISK	TAPE	CPU	DISK	TAPE
ALICE	126000	14300	33170	126000	14300	33170
ATLAS	136440	14670	40680	136440	14670	40680
CMS	120900	15680	49400	120900	15680	49400
LHCB	113430	11561	25261	113430	11561	25261
LHC TIER1	**496770**	**56211**	**148511**	**496770**	**56211**	**148511**
AMBER	0	0	0	1000	50	0
Belle2	31000	1320	650	31000	1320	710
CDF	0	0	4000	0	0	4000
DUNE_CSN1	5000	1100	510	5000	1100	1010
FCC	1000	200	0	2000	200	0
GMINUS	640	160	1200	640	160	1200
HYPERK_CSN1	0	0	0	10838	152	905
ICAR-US	22000	1600	4000	5000	2000	5000
KLOE	0	33	3075	1000	333	3075
LHCB TIER2	62600	0	0	87500	0	0
LHCf	12000	170	0	18000	170	50
Mu2e	0	0	0	0	0	50
MuonCollider	3000	150	150	0	150	150
MUONE	1000	100	650	1000	100	650
NA62	3300	275	3300	3300	275	3300
PADME	4417	100	1780	4417	100	1780
Gruppo 1	**145957**	**5208**	**19315**	**170695**	**6110**	**21880**
AMS2	31833	2800	1650	35833	2800	2000
AUGER	5430	1100	300	8430	1400	1600
Borexino	500	359	80	500	359	80
BULLKID	125	10	0	125	10	0
CTA	5296	2000	2700	5296	2000	3700
CUORE	3000	850	100	3000	950	200
CUPID	905	25	10	905	25	10
CYGNO	417	40	200	417	40	200
DAMPE	27306	850	10	27306	900	10
DARKSIDE	6000	3150	1920	6500	3150	1920
ENUBET	250	10	5	250	10	5
ET	500	55	0	500	55	0
EUCLID	0	1450	1000	400	1450	1000
FERMI-GLAST	350	15	20	350	15	20
GAPS	1863	80	0	2233	90	0
Gerda	40	50	50	40	50	50
Herd	8111	450	0	8111	450	0
hyperk	10838	152	605	0	0	0
JUNO	19983	3000	1000	19983	6000	4000
KM3	7500	450	250	7500	450	250
LIMADOU	2300	180	12	4000	260	12
MAGIC	0	0	30	0	0	30
NEWS	284	300	150	284	300	150
NUCLEUS	500	170	83	1000	170	83
PAMELA	747	110	150	747	110	150
QUAX	0	0	120	0	0	120
SPB2_MiniEUSO	200	10	0	200	20	0
SWGO	400	300	0	400	500	0
Tristan	2000	40	0	3000	60	0
Virgo	50000	900	5158	50000	944	6148
Xenon100	1250	600	3500	1250	800	3500
Gruppo 2	**187928**	**19506**	**19103**	**188560**	**23368**	**25238**
Agata-GAMMA	2500	254	1860	2500	354	1910
ASFIN	1042	212	0	1042	212	0
EIC-NET	950	100	0	950	100	0
FAMU	2932	43	15	2932	53	35
FOOT	914	120	0	1014	120	0
JLAB12	12167	150	0	12167	150	0
LUNA	500	10	50	500	10	50
NEWCHIM-FARCOS	190	200	860	190	200	860
n-TOF	9890	15	0	9890	15	0
Nuclex-Fazia	0	50	0	0	50	0
Gruppo 3	**31085**	**1154**	**2785**	**31185**	**1264**	**2855**
Gruppi 1+2+3	**364970**	**25868**	**41203**	**390440**	**30742**	**49973**
Total	**861740**	**82079**	**189714**	**887210**	**86953**	**198484**
Overlap	**843740**	**82079**	**189714**	**887210**	**86953**	**198484**

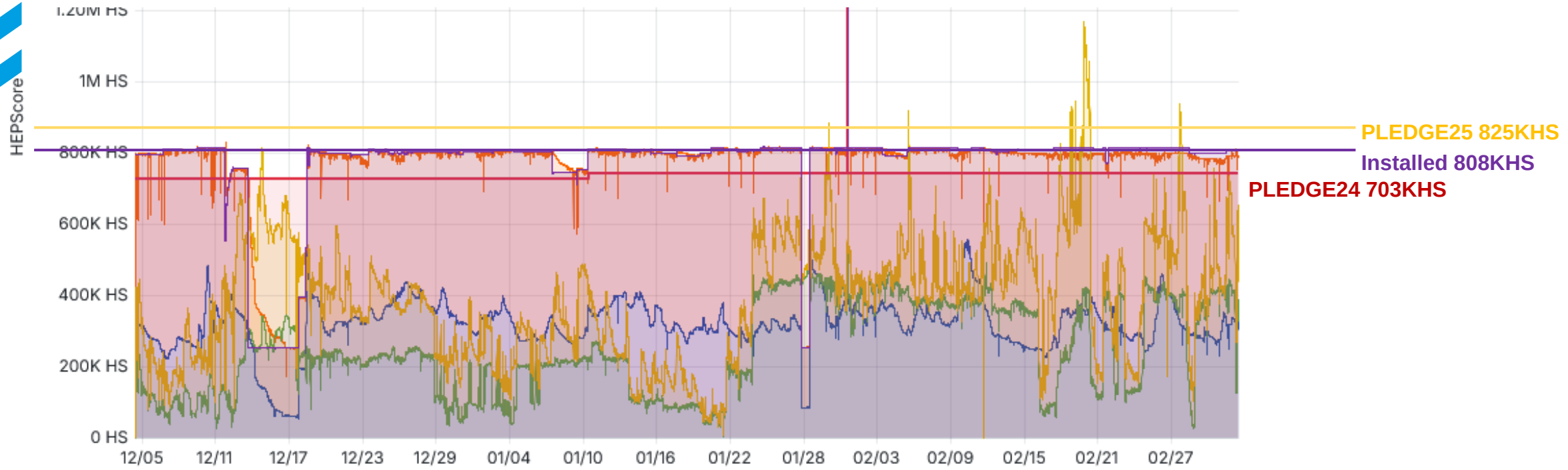
Resources PLEDGE @T1 2024-2025

ALL VO No Cloud	2024	2025	Delta
Pledge CPU (HS06)	703000	792000	89000
Pledge disk (TBN)	82079	101023	18944
Pledge tape (TB)	190214	233374	43160

CPU Farm at T1

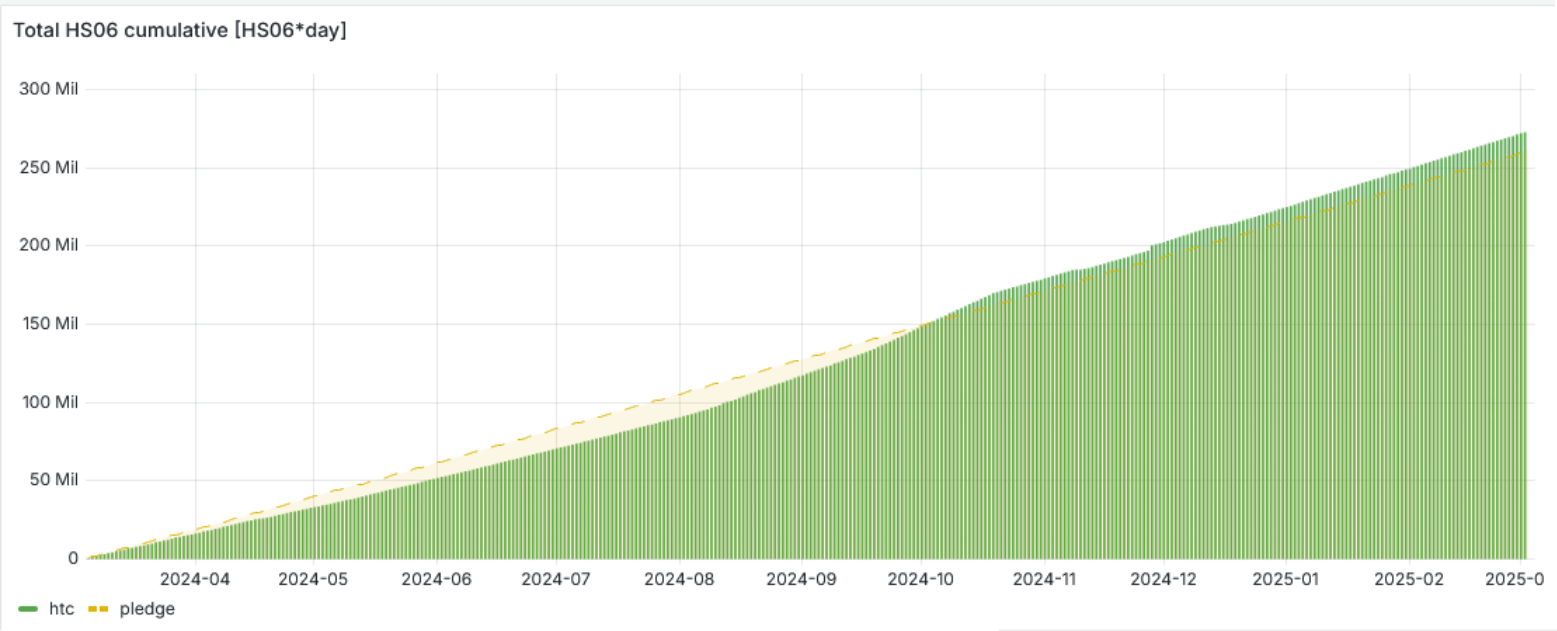
Pledge 2024: 703kHS06 ➔ **Potenza installata Totale: 808KHS06**

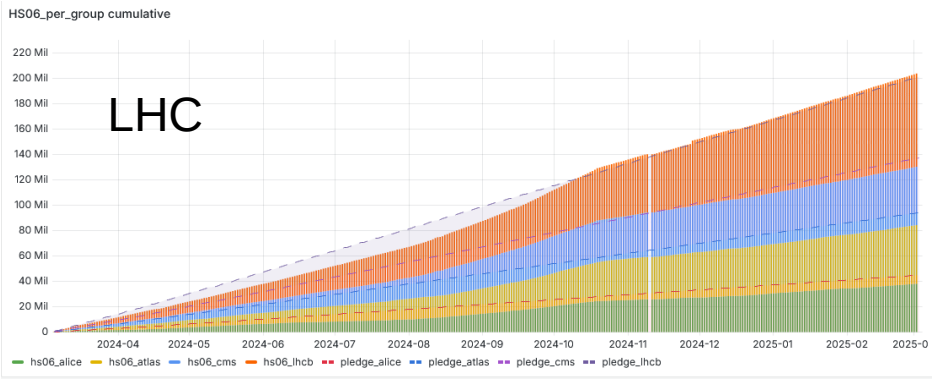
Pledge2025: 825KHS

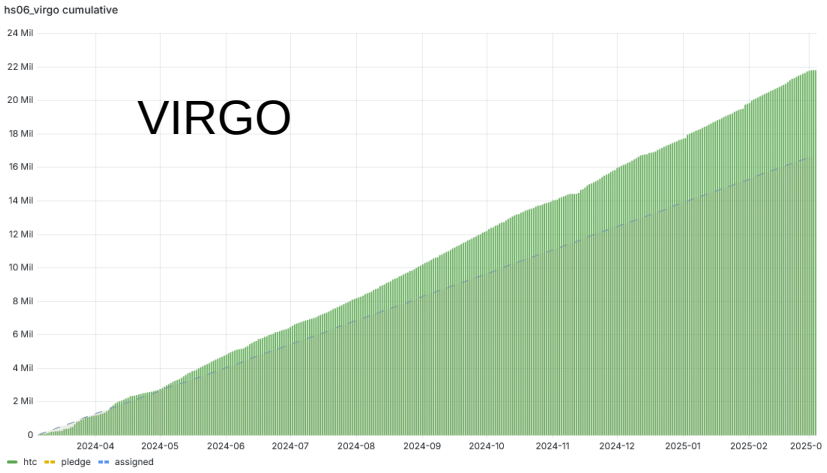


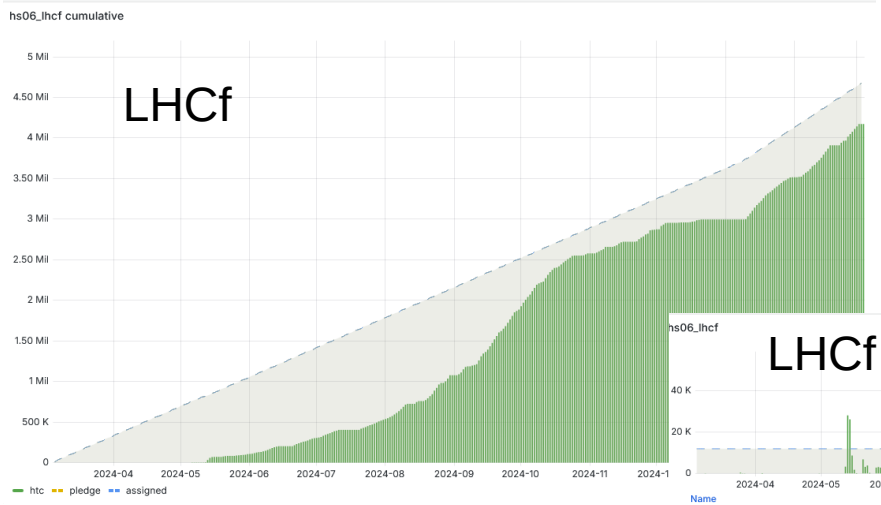
Name	Last *	Mean
running - multi_core	305K HS	319K HS
running - single_core	488K HS	451K HS
idle - multi_core	327K HS	251K HS
idle - single_core	113K HS	138K HS
pledge	743K HS	736K HS
available	808K HS	772K HS

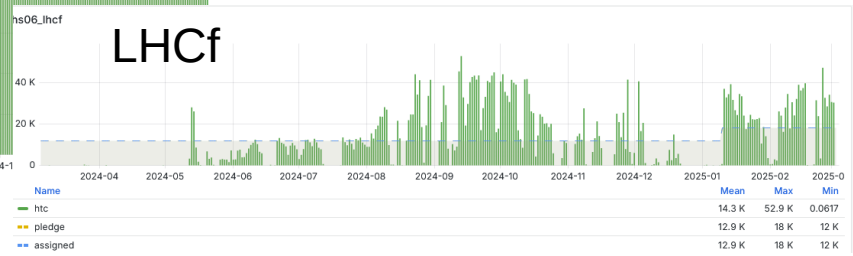
HS06 integrati





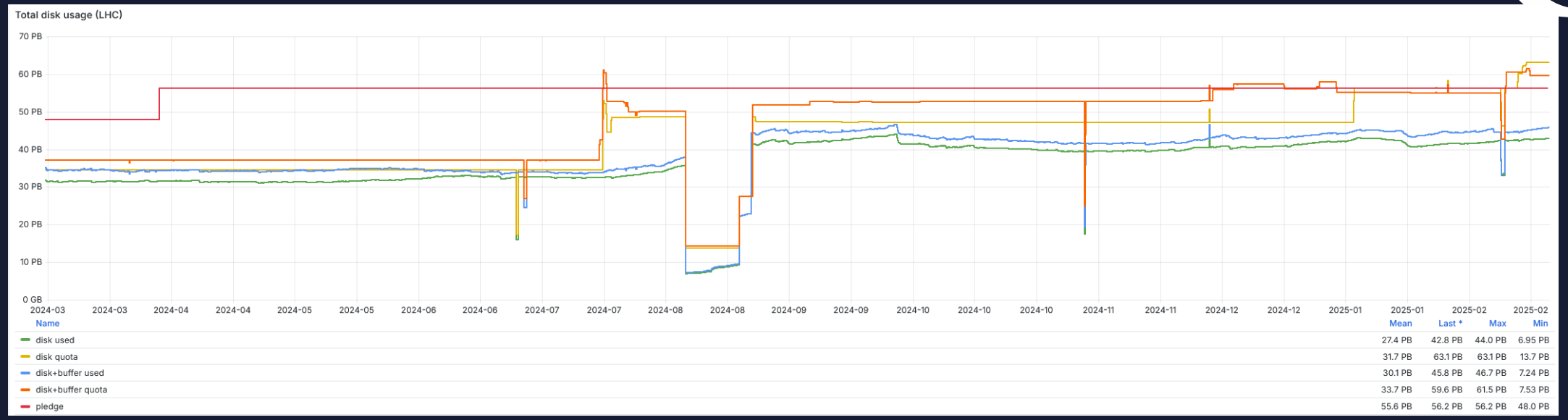


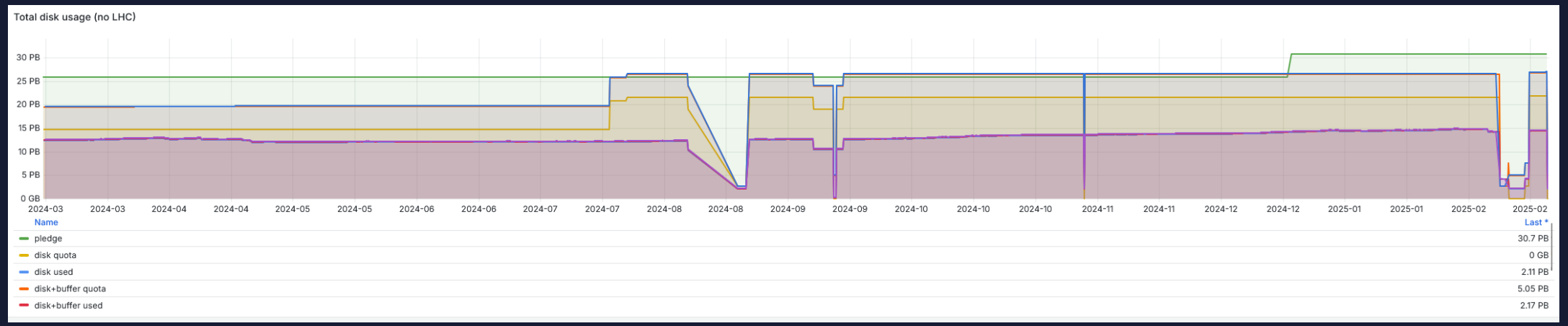




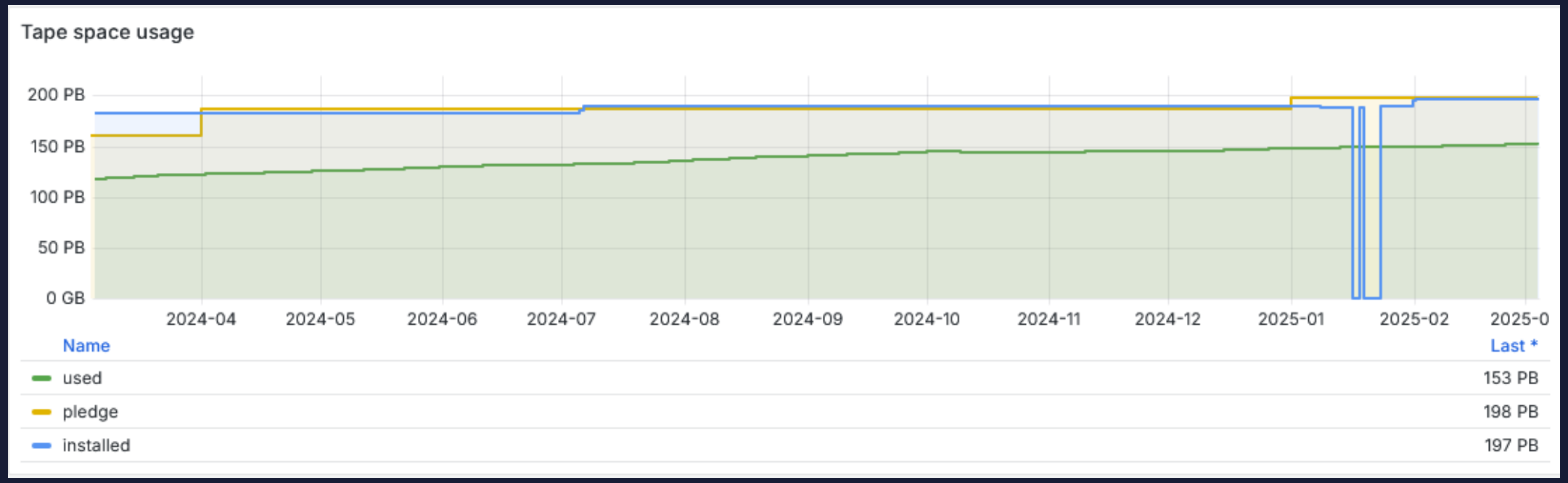
DISK and TAPE al T1

DISK USAGE @T1

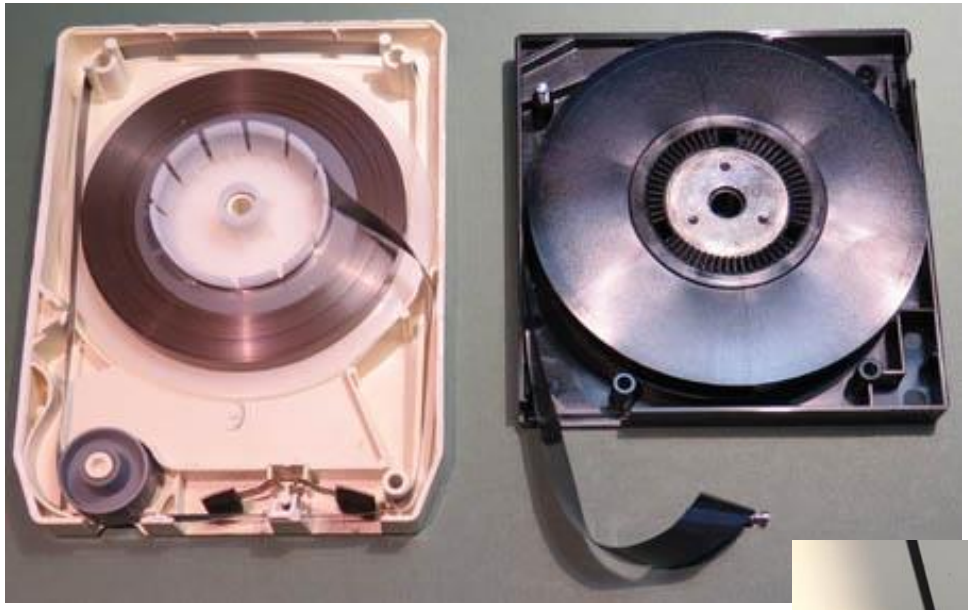




TAPE@T1


Tape space usage
200 PB
150 PB
100 PB
50 PB
0 GB
2024-04
2024-05
2024-06
2024-07
2024-08
2024-09
2024-10
2024-11
2024-12
2025-01
2025-02
2025-0
Name
Last *
used
153 PB
pledge
198 PB
installed
197 PB

TAPE devices



Back to the 80s

Commodore’s datassette: a 90-minutes tape (45 minutes on each side) will hold on the order of 150 kilobytes on each side if no compression or fast loader is used.

Back to the 80s

Zak McKracken and the Alien Mindbenders
Released: 1988 (36 years ago)
Publisher: Lucasfilm GamesInfo / Logos
Coder: David Fox
Matthew Kane
Graphics: Gary Winnick
Martin Cameron
Musician: Matthew Kane
Sound FX: Chris GriggInfo
Matthew Kane
Box Art: Steve Purcell

Bubble Bobble

Published: 1987, Firebird
Category: Platformer - Single Screen
Players: 1 or 2, Simultaneous

Turrican II: The Final Fight
Published: 1991, Rainbow Arts
Category: Shoot'em Up - Platformer
Players: 1 Only

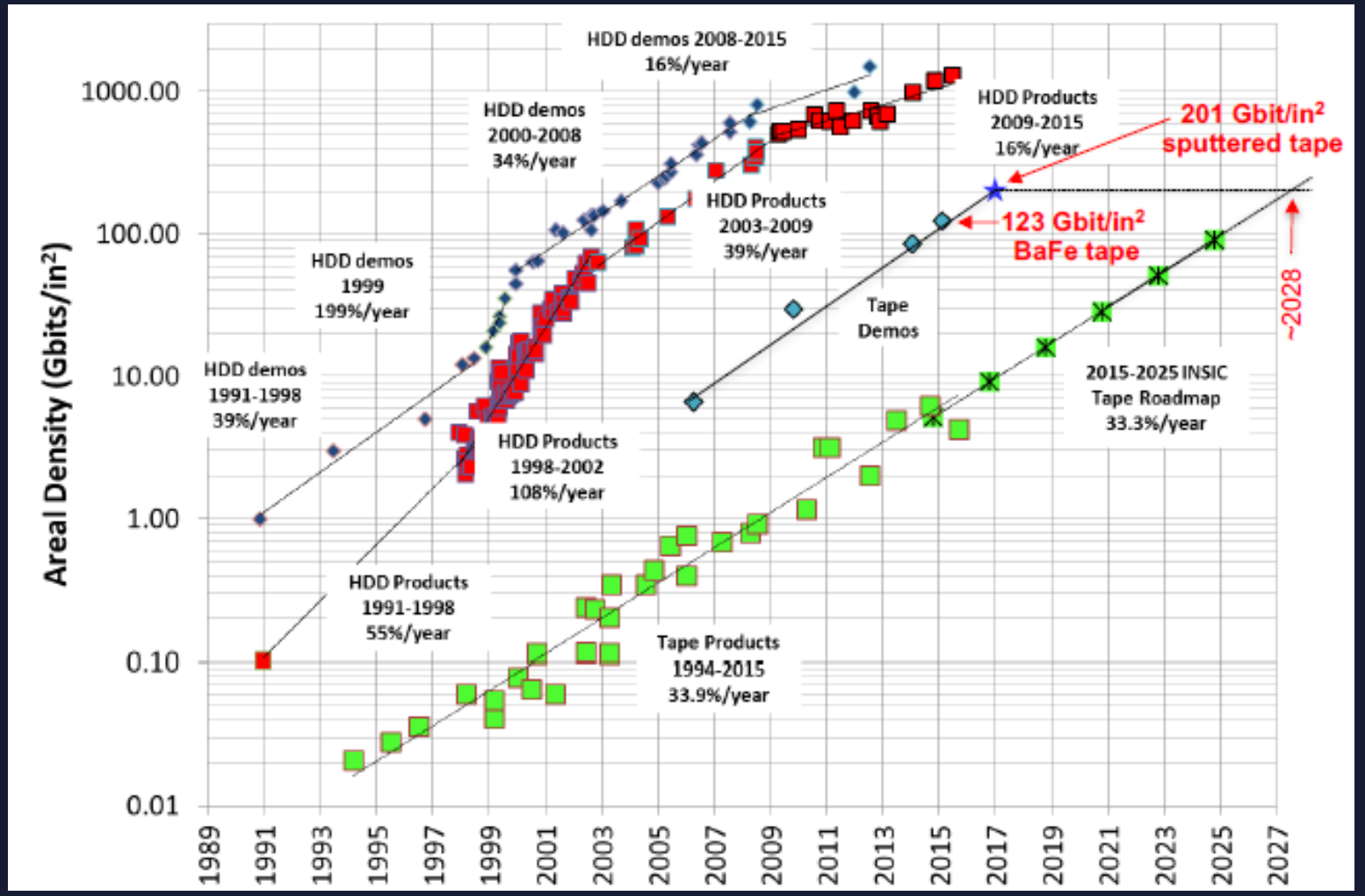
Tape area network

- Is the part of the SAN dedicated to the interconnection among servers, libraries and tape drives
- Tape drives can be installed in a central array and attached to the SAN, making them accessible to every server on the network

Tape Libraries at CNAF

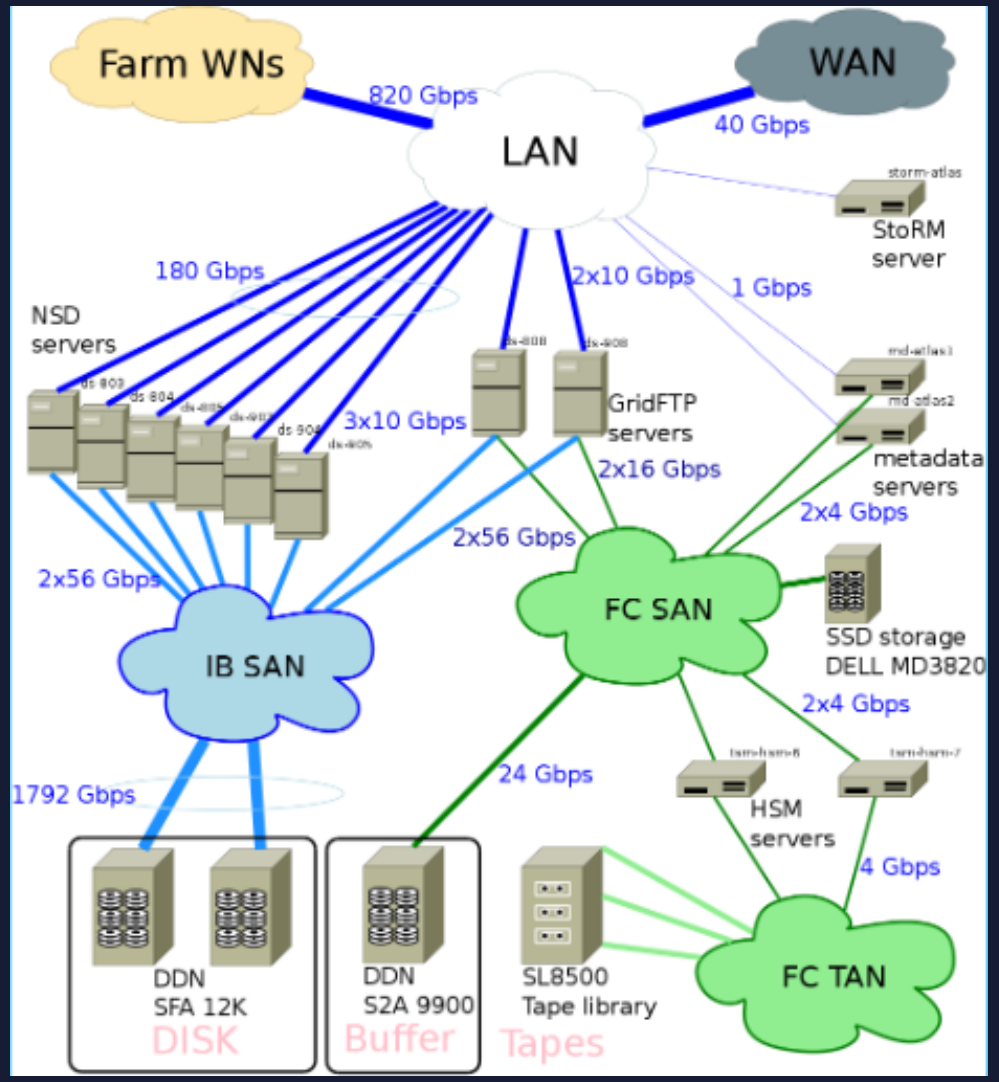
Library	Tape drives	Max data rate/drive, MB/s	Max slots	Max tape capacity, TB	Installed cartridges	Used space, PB	Free space, PB
SL8500 (Oracle)	16*T10KD	250	10000	8.4	~10000	**47**	32
TS4500 (IBM)	19*TS1160	400	6198	20	5100+380	**102**	**0.6**
TS4500-2(IBM)	18*TS1170	400	7844	50	165	**0**	**8.2**

Areal density scaling



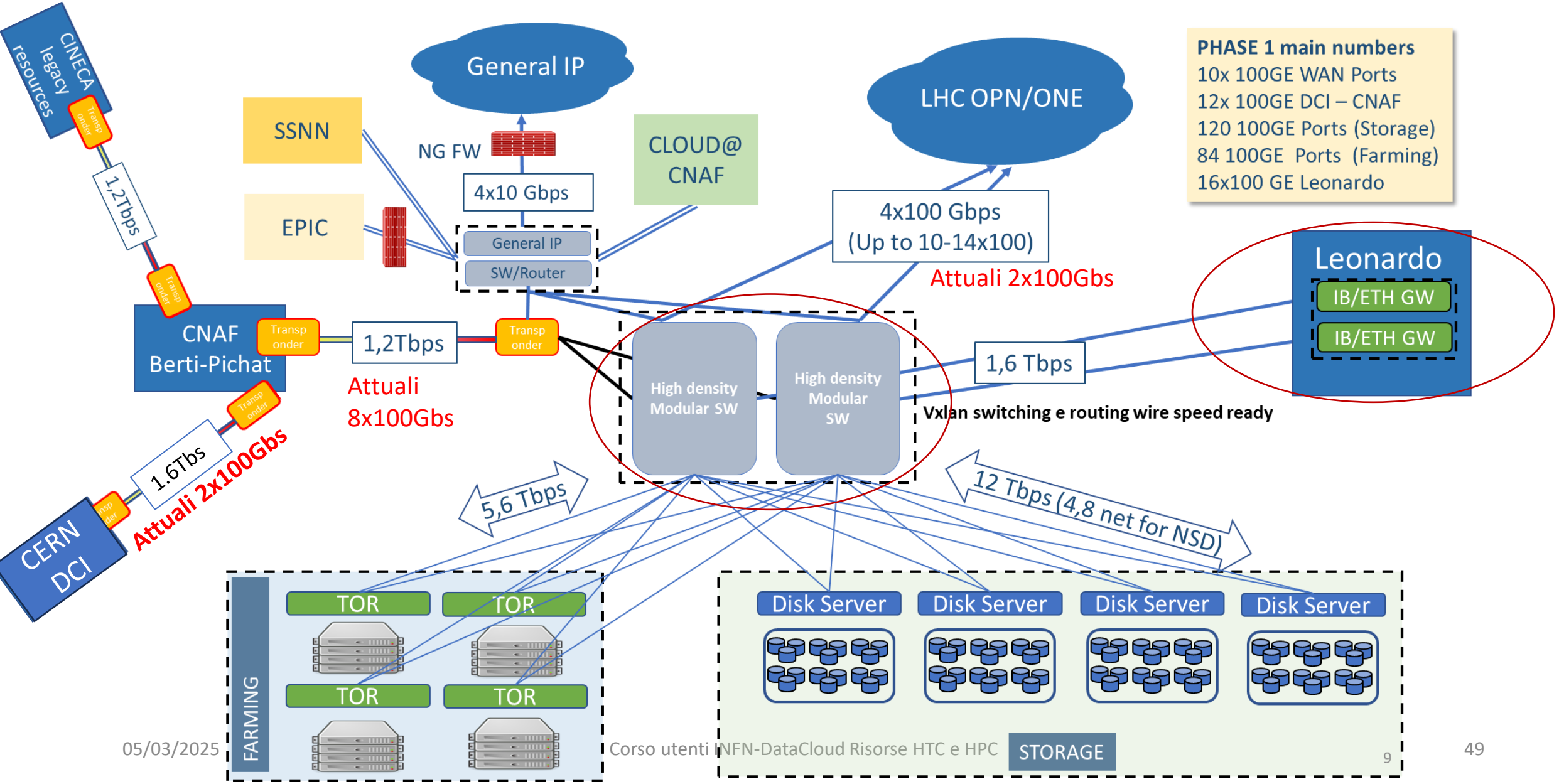
- 2015: IBM-FujiFilm demonstration of 123 Gb/in2 on BaFe tape
- 2017: IBM-Sony demonstration of 201 Gb/in2 on Sputtered Tape

A Storage Area Network at CNAF

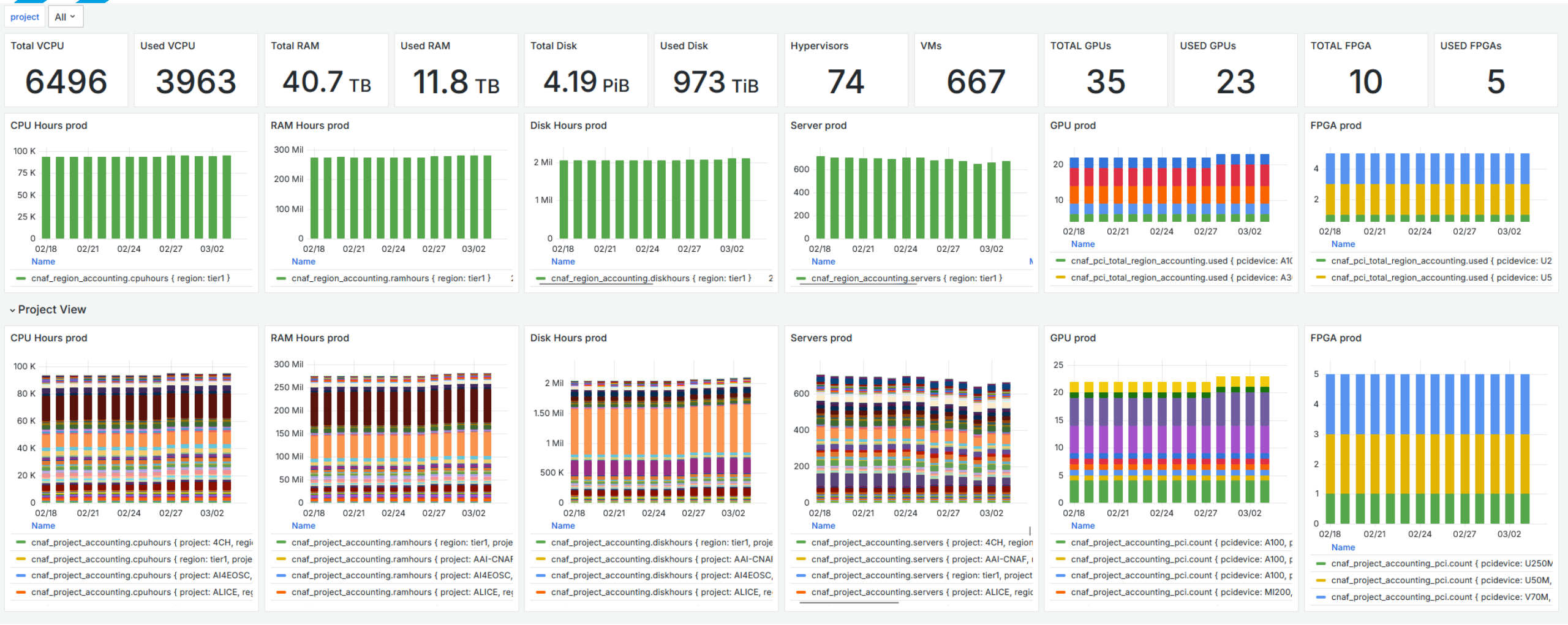


- 7 PB disk space
- 6 NSD servers (3x10 Gbps)
- 2 metadata servers (1Gbps)
- 2 GridFTP (XrootD) (2x10 Gbps)
- 2 HSM servers
- Metadata on SSD (mirrored)
- VM as Storm server
- Throughput required (5 MB/s/TB) = 21.5 GB/s
- Throughput available (6 NSD x 30 Gbps) = 22.5 GB/s

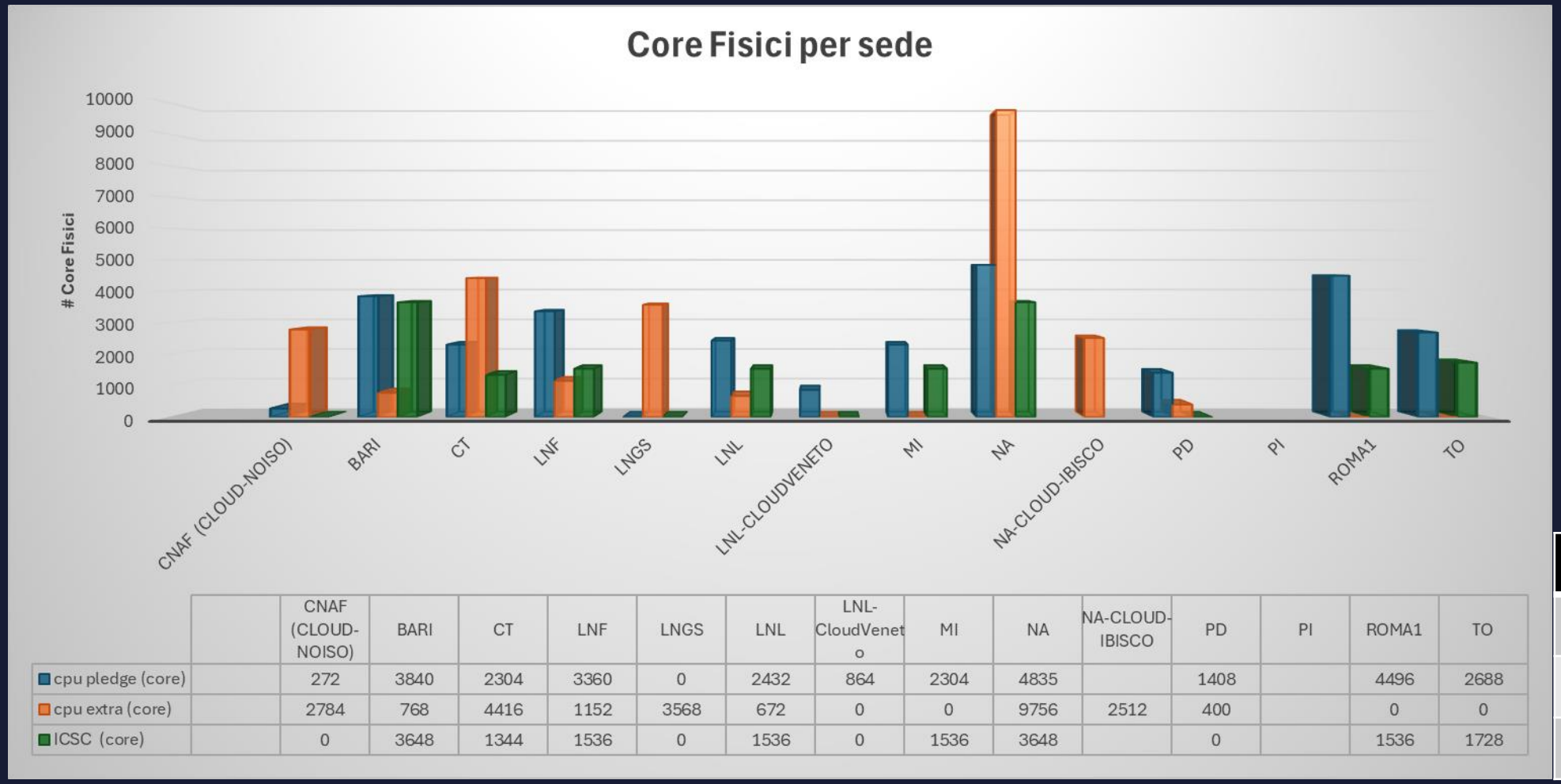
Networking Infrastructure at CNAF



Cloud Resources at CNAF



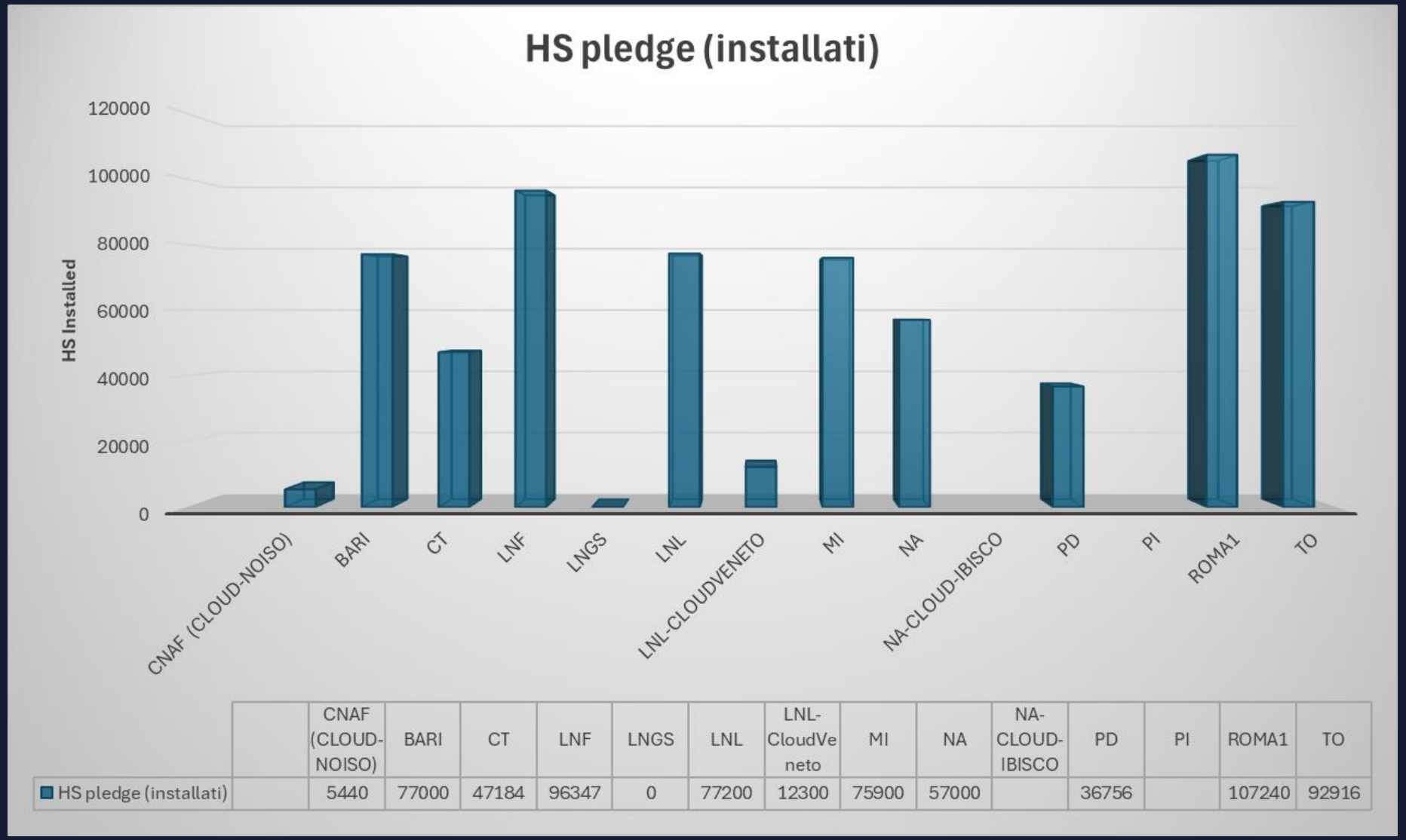
Risorse ai Tier-2



		CNAF (CLOUD-NOISO)	BARI	CT	LNF	LNGS	LNL	LNL-CloudVeneto	MI	NA	NA-CLOUD-IBISCO	PD	PI	ROMA1	TO
cpu pledge (core)		272	3840	2304	3360	0	2432	864	2304	4835		1408		4496	2688
cpu extra (core)		2784	768	4416	1152	3568	672	0	0	9756	2512	400		0	0
ICSC (core)		0	3648	1344	1536	0	1536	0	1536	3648		0		1536	1728

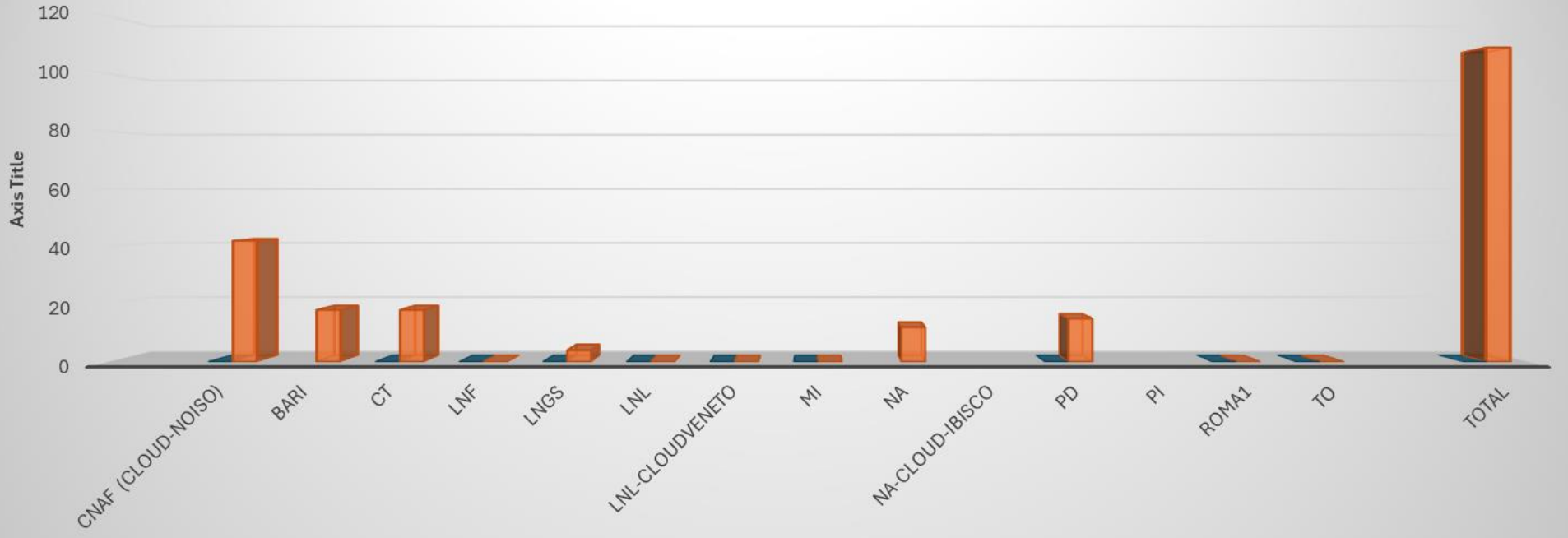
	TOTAL
Pledged	28803
Extra	26028
ICSC	16512

ICSC expected = 18048
1728 da PISA

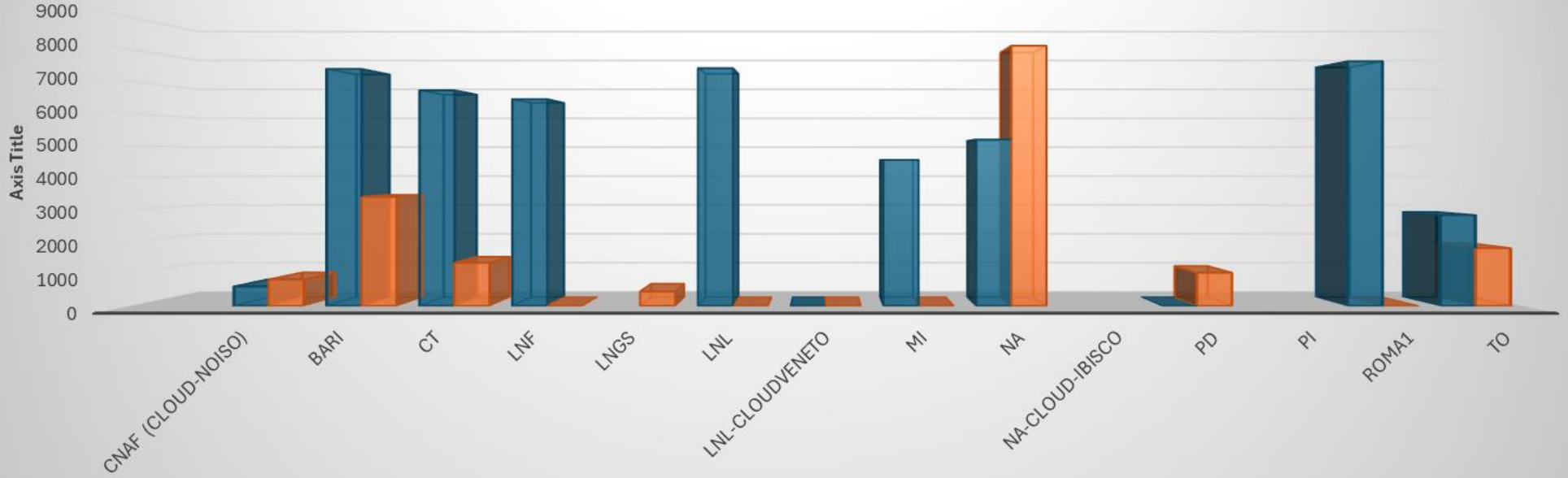


	CNAF (CLOUD-NOISO)	BARI	CT	LNF	LNGS	LNL	LNL-CloudVeneto	MI	NA	NA-CLOUD-IBISCO	PD	PI	ROMA1	TO
HS pledge (installati)	5440	77000	47184	96347	0	77200	12300	75900	57000		36756		107240	92916

TOTAL=685283
CRIC(WLCG T2)= 567275



		CNAF (CLOUD-NOISO)	BARI	CT	LNF	LNGS	LNL	LNL-CloudVeneto	MI	NA	NA-CLOUD-IBISCO	PD	PI	ROMA1	TO		TOTAL
gpu pledge		0		0	0	0	0	0	0			0		0	0		0
gpu extra		42	18	18	0	4	0	0	0	12		15		0	0		109



		CNAF (CLOUD-NOISO)	BARI	CT	LNF	LNGS	LNL	LNL-CloudVeneto	MI	NA	NA-CLOUD-IBISCO	PD	PI	ROMA1	TO
disk pledge (TB-N)		605	7387	6720	6444		7412	0	4554	5190		0		7624	2830
disk extra (TB-N)		825	3400	1344	0	450	0	0	0	8110		1037		0	1800

	TOTAL TB-N
Pledged	48766
Extra	16966

CRIC Total Pledge T2 (WLCG): 4660TB-N

Risorse della seconda tornata di acquisti (2024)

Tier-2, esclusi sistemi HPC

(i.e. solo quota ICSC)

Potenza CPU:

~17 HS06/coreHT

→ ~287 kHS06

Storage disco (gara dedicata da effettuare):

Tradizionale: TBN = ~0.73*TBL

CEPH: TBN = ~0,67*TBL

→ ~50 PBN

16PBL da gara HPC bubble

Incluso potenziamento INFNCLOUD backbone a CNAF e BARI

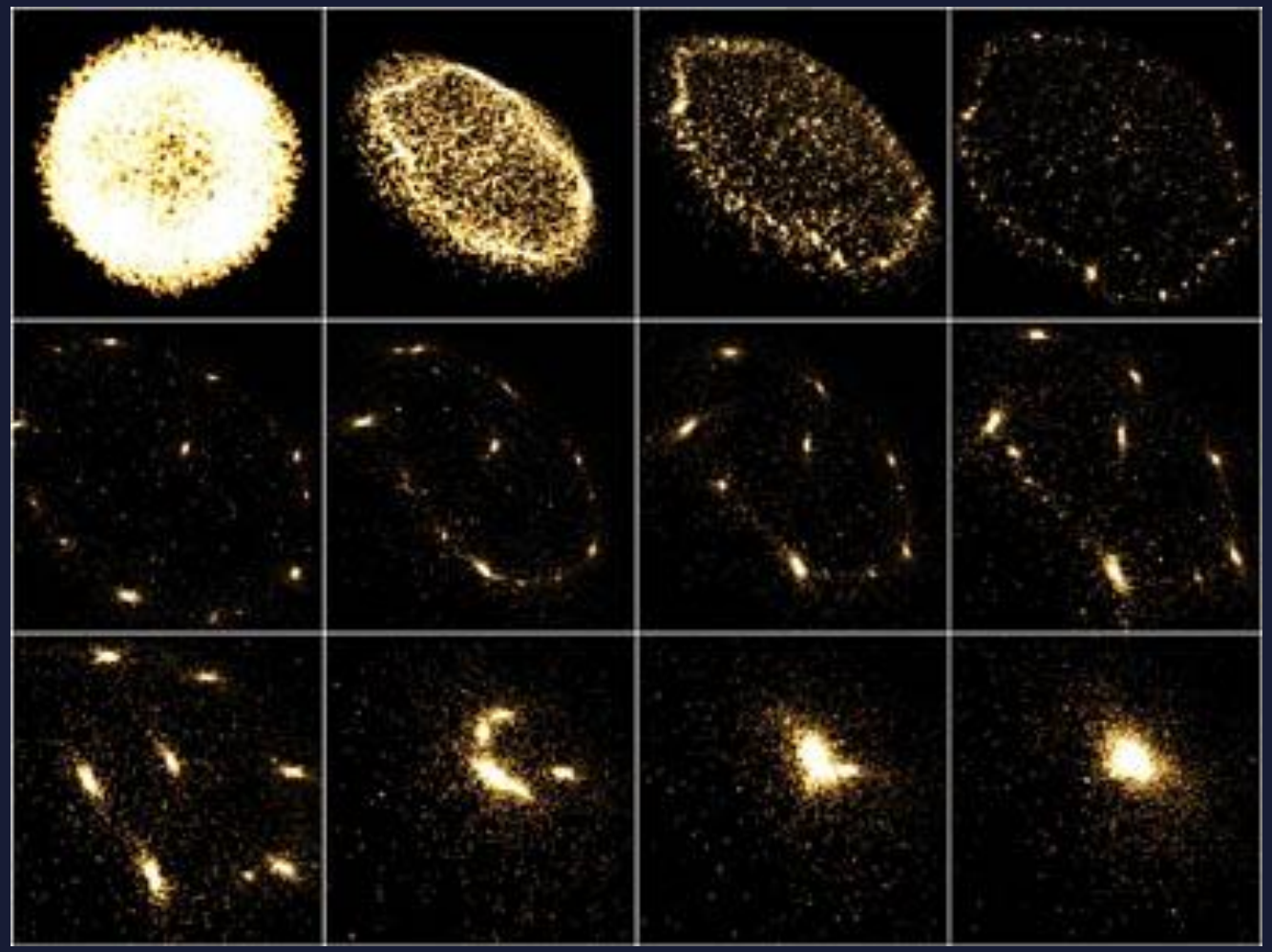
Sito	CPU (Core fisici)	Storage (PBL)
BA	2304	12.2
CT	2304	16.7
LNF	2304	2.5
LNFESA	1536	2.3
LNGS	-	4.6
LNL	784	5.8
PD	-	
MI	784	3.9
NA	2304	12.2
RM1	784	4.5
PI	2304	3.2
TO	1536	4.5
CNAF	-	-
TOT	16896	72.4

High Performance Computing
HPC

HTC and HPC - definition

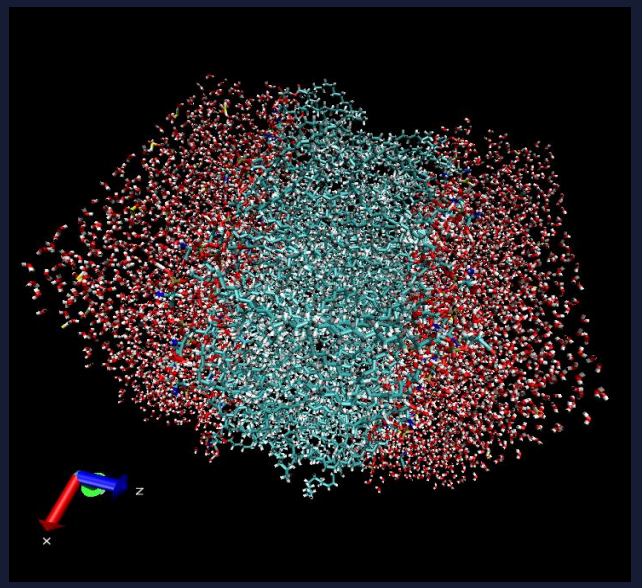
- High Throughput Computing (HTC)
  - The focus is on the execution of many copies of the *same program* at the *same time*
    - not in the speedup of individual jobs
  - Many copies of the same program run *in parallel* or *concurrently*
  - Maximize the **throughput**
- High Performance Computing (HPC)
  - speed up the individual job as much possible so that results are achieved more quickly
- HTC infrastructures tend to deliver large amounts of computational power over a long period of time.
  - In contrast, High Performance Computing (HPC) environments deliver a tremendous amount of compute power over a short period of time.
- The interest in HTC is in how many jobs complete over a long period of time instead of how fast an individual job can complete.

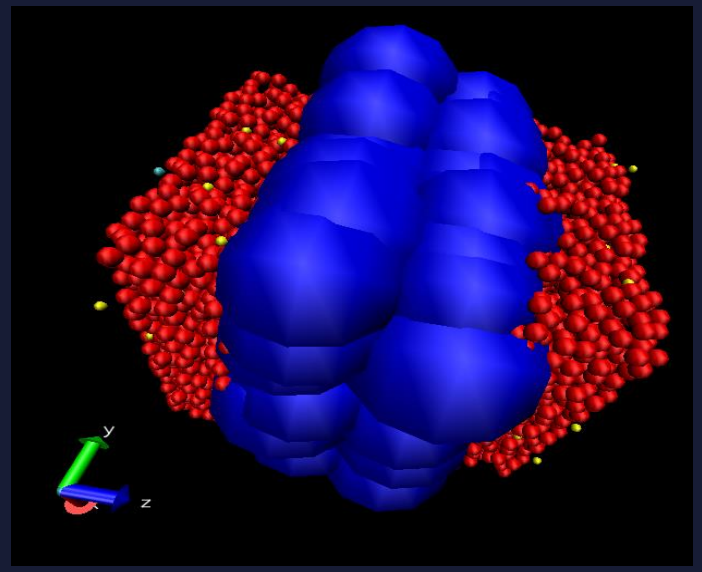
HPC Applications



HPC - Applications

- Molecular Dynamics

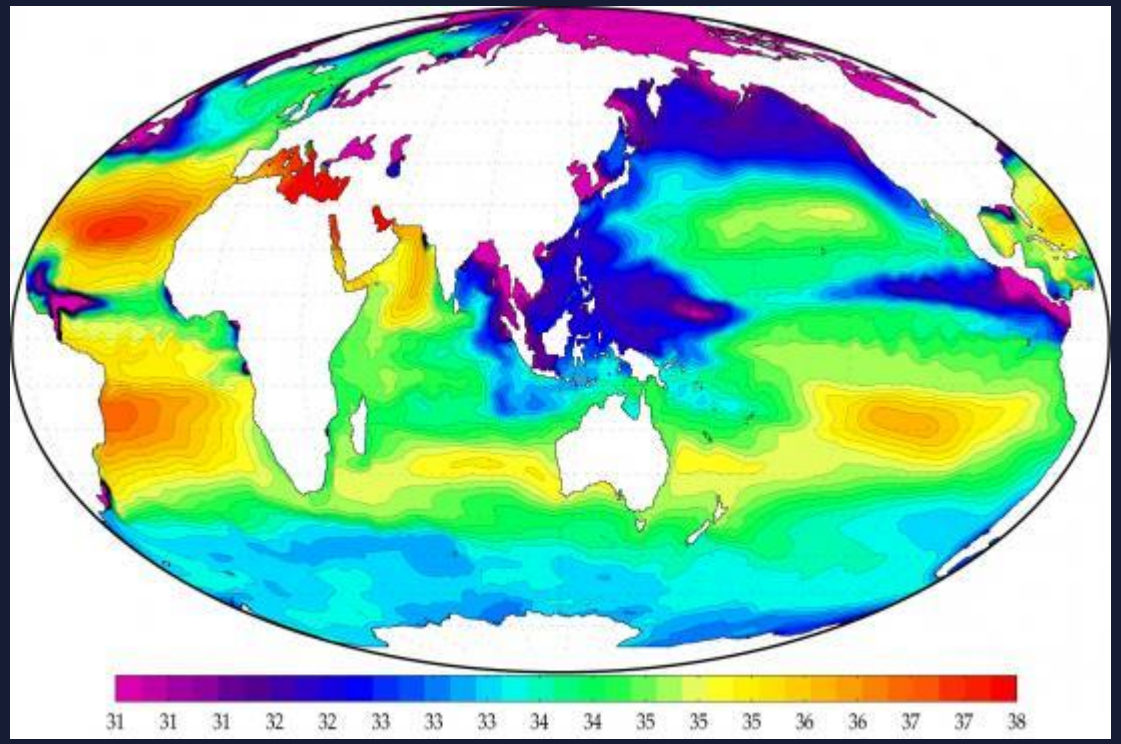


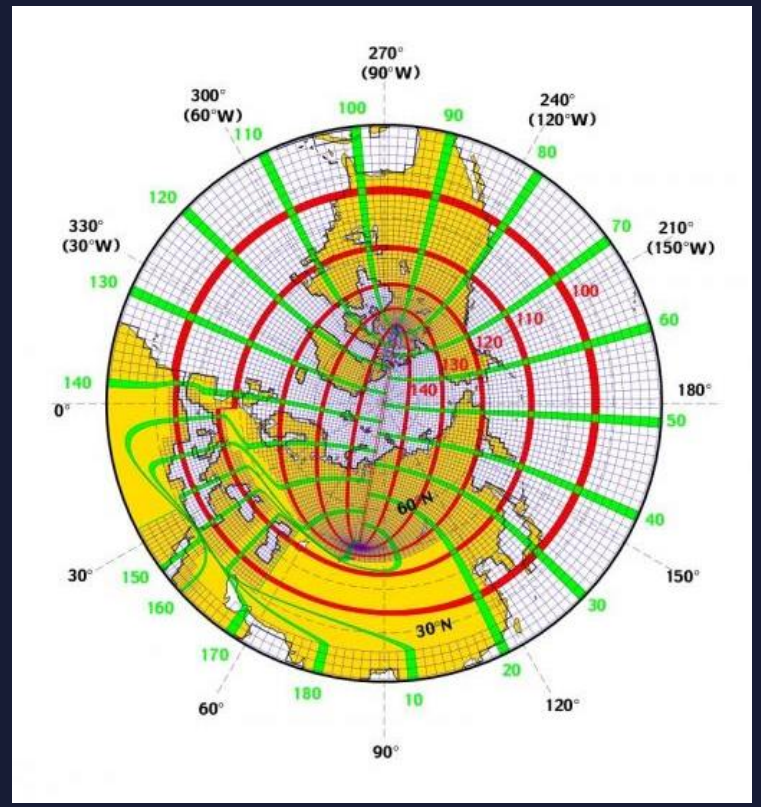


NAMD, Quantum Espresso, Gromacs, Gaussian, etc..

HPC - Applications

- Earth simulation

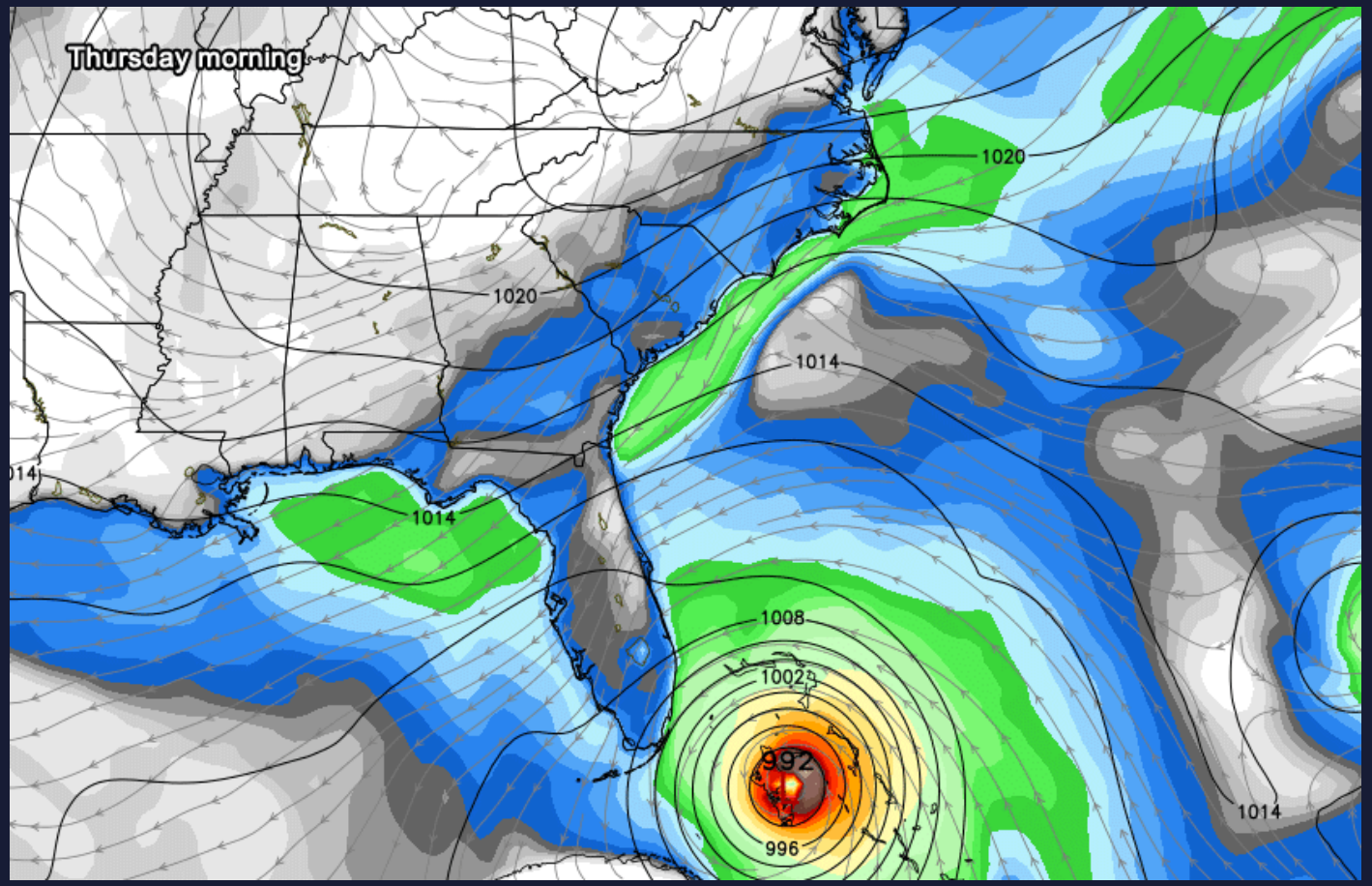




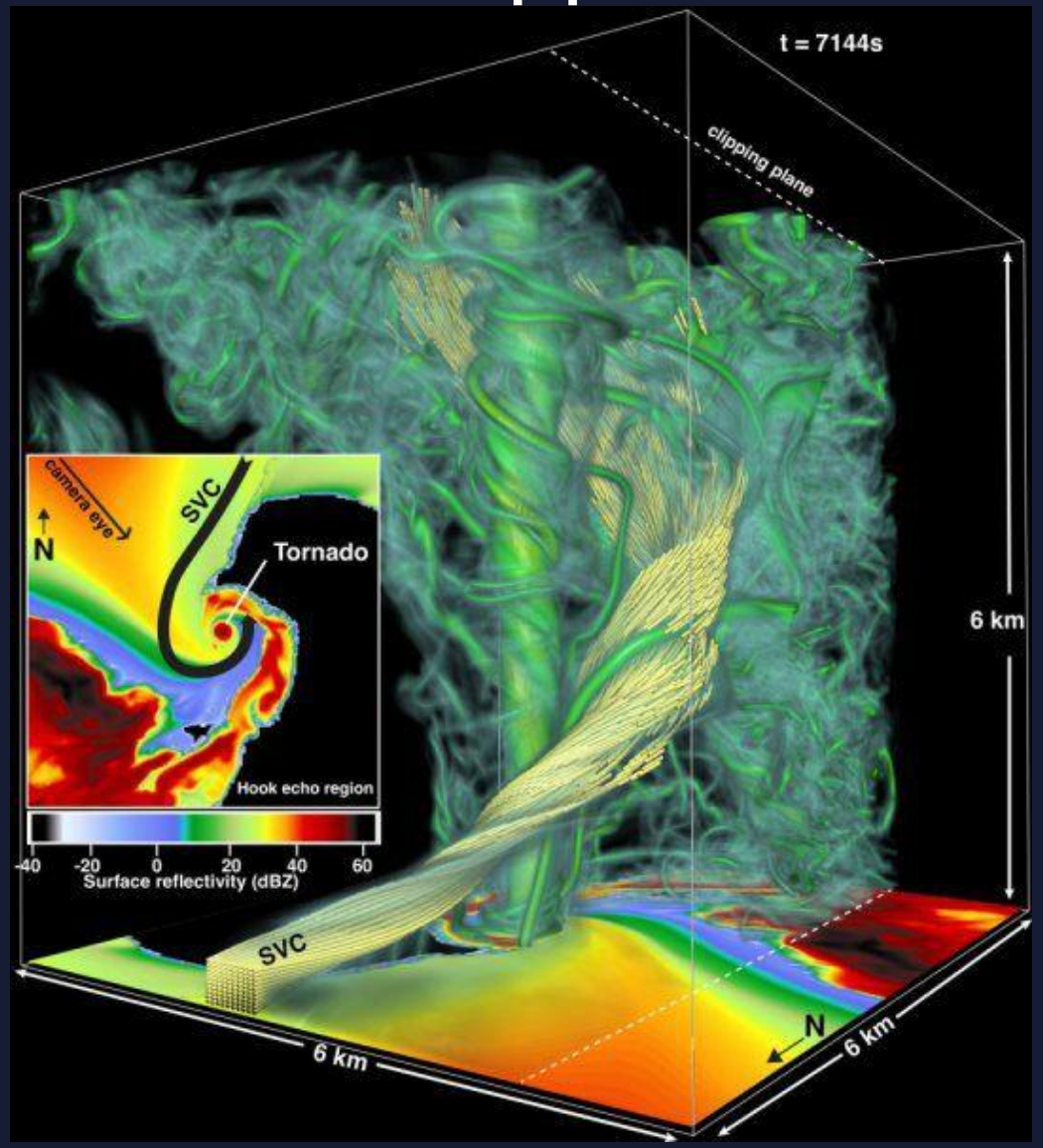
WRF, MM5, GLOBO, NEMO, etc..

HPC- Applications

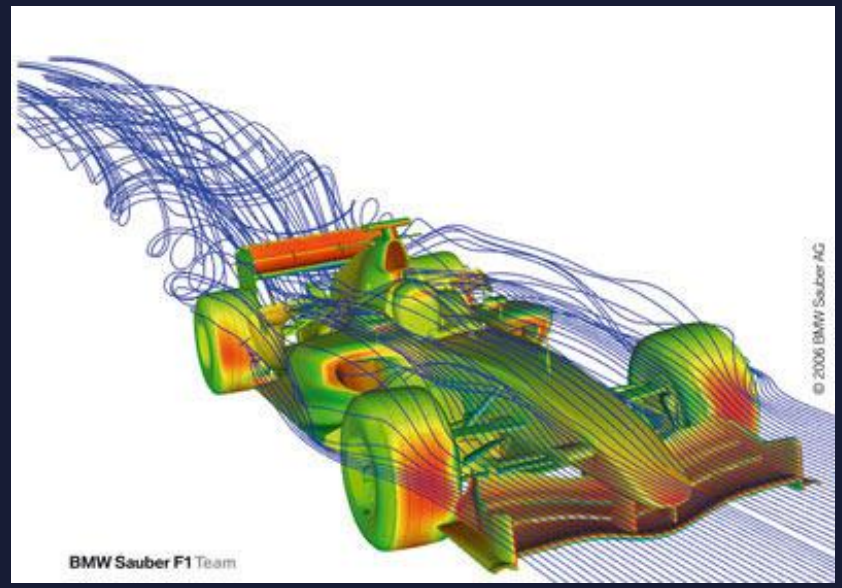
- Weather Simulation



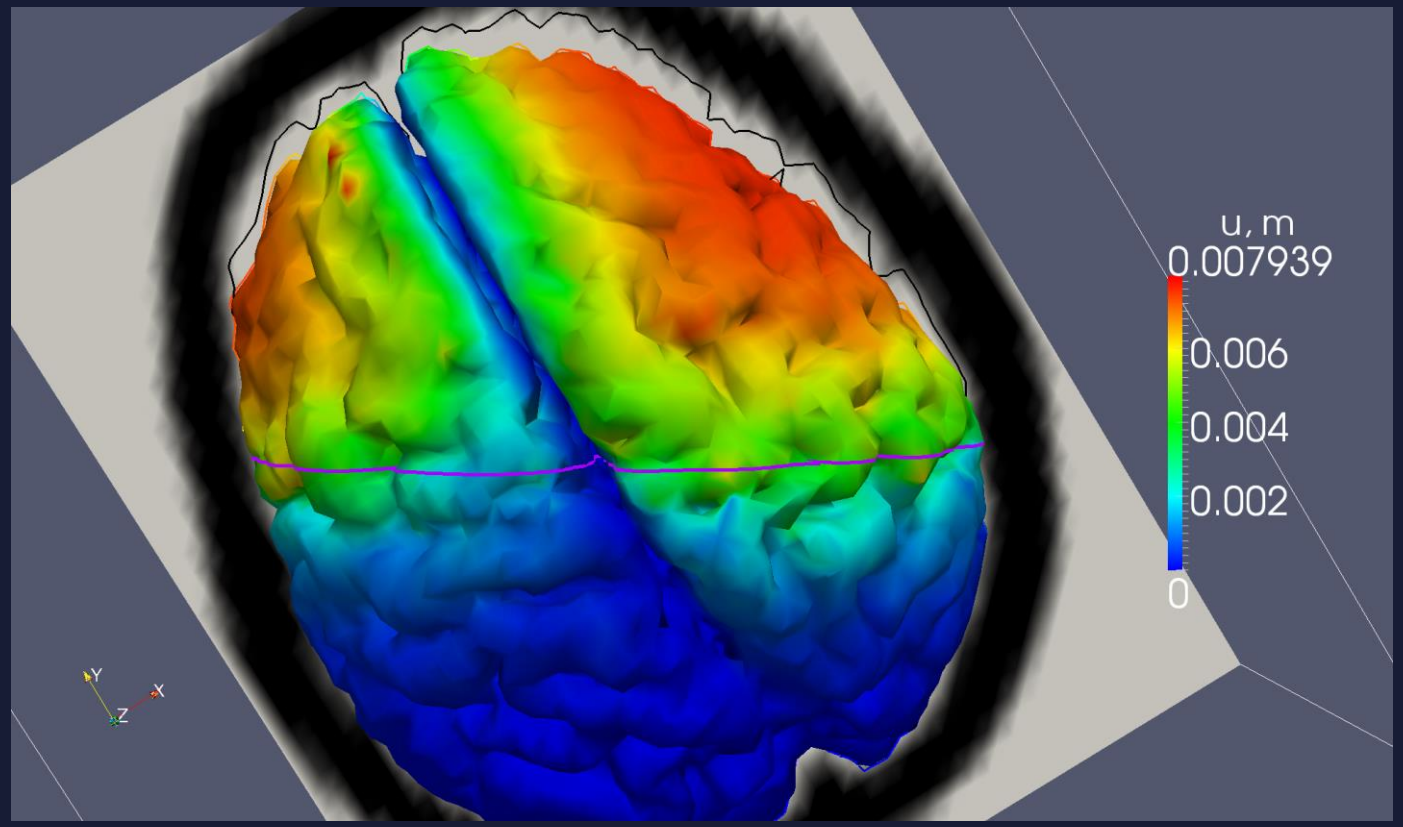
HPC - Applications



- Fluid Dynamics



HPC - Applications



- Brain Simulation

Once upon a time....

The vector machines

- Serial number 001 Cray-1™
  - Los Alamos National Laboratory in 1976
  - $8.8 million
  - 80 MFLOPS scalar, 160/250 MFLOPS vector
  - 1 Mword (64 bit) main memory
  - 8 vector registers
    - 64 elements 64bit each
  - Freon refrigerated
  - 5.5 tons including the Freon refrigeration
  - 115 kW of power
    - 330 kW with refrigeration

(*) Source: Wikipedia

Serial number 003 was installed at the National Center for Atmospheric Research (NCAR) in 1977 and decommissioned in 1989

El Capitain

- Hewlett Packard Enterprise El Capitan is an exascale supercomputer
- Hosted at the Lawrence Livermore National Laboratory in Livermore, California, United States, became operational in 2024.
- It is based on the Cray EX Shasta architecture. El Capitan displaced Frontier as the world's fastest supercomputer in the 64th edition of the Top500 (Nov 2024).
- Its primary purpose is to support the stockpile stewardship program of the US National Nuclear Security Administration
- Uses a combined 11,039,616 CPU and GPU cores consisting of 43,808 AMD 4th Gen EPYC 24C "Genoa" 24-core 1.8 GHz CPUs (1,051,392 cores) and 43,808 AMD Instinct MI300A GPUs (9,988,224 cores).

Active	Deployment: 2H 2023 Completion: 2024
Sponsors	U.S. Department of Energy
Operators	Lawrence Livermore National Laboratory and U.S. Department of Energy
Location	Livermore Computing Complex
Architecture	HPE Cray Shasta
Power	30 MW[1]
Operating system	TOSS
Space	TBA
Memory	5.4375 petabytes[2]
Storage	TBA
Speed	1.742 exaFLOPS (Rmax) / 2.746 exaFLOPS (Rpeak)[2]
Cost	US$600 million (estimated cost)
Purpose	Scientific research and development, stockpile stewardship[3]

© Wikipedia

LEONARDO@CINECA

- **Petascale supercomputer located at the CINECA datacenter in Bologna, Italy.**
- **Atos BullSequana XH2000 computer**
  - **14,000 Nvidia Ampere GPUs**
  - **200 Gbit/s Nvidia Mellanox HDR InfiniBand connectivity.**
- **250 petaflops**
  - **top five in TOP500**
  - **second in Europe**

Leonardo

Active	November 24, 2022
Sponsors	European High-Performance Computing Joint Undertaking
Operators	CINECA
Location	Bologna, Italy
Architecture	13,824 Nvidia Ampere GPU cores
Power	6 MW
Space	900+ m^2
Memory	2.8 petabytes
Storage	110 petabytes
Speed	250 petaFLOPS (peak)
Cost	€240 million
Website	Leonardo Pre-exascale Supercomputer

© Wikipedia

LEONARDO@CINECA - 2

- **Booster Module**
  - The 3,456 individual nodes which make up the "booster module" are custom BullSequana X2135 "Da Vinci" blade servers, each composed of:
    - 1x Intel Xeon 8358 CPU, with 32 cores running at 2.6 GH
    - 512 GB RAM DDR4 3200 MHz
    - 4x NVidia custom Ampere GPU, 64 GB HBM2
    - 2x NVidia HDR InfiniBand network adapters, each with two 100 Gbit/s ports
    - Each node is expected to deliver 89.4 TFLOPs peak.
- **Data Centric Module**
  - The "data centric module" consists of 1536 nodes, each comprising a BullSequana X2610 compute blade with:
    - 2x Intel Sapphire Rapids CPUs, with 56 cores
    - 512 GB RAM DDR5 4800 MHz
    - 1x NVidia HDR InfiniBand network adapter, with one 100 Gbit/s port
    - 8 TB NVM storage

Clusters

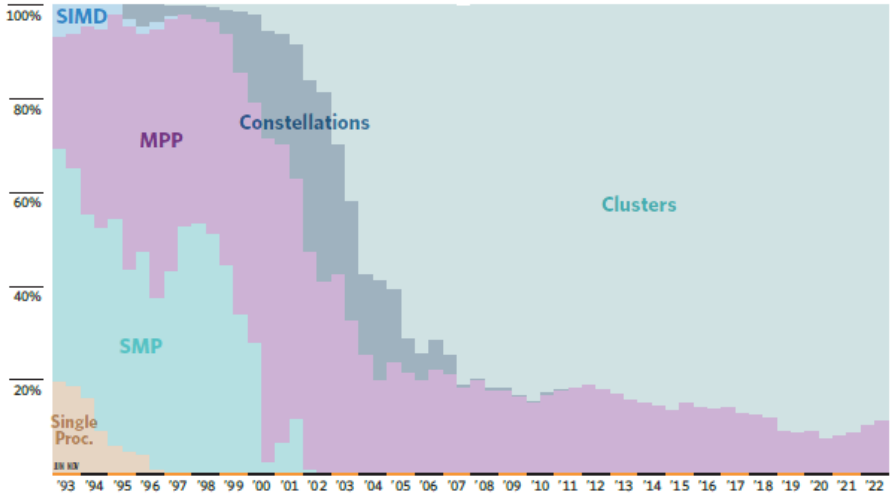
[a cluster is a] parallel computer system comprising an integrated collection of independent nodes, each of which is a system in its own right, capable of independent operation and derived from products developed and marketed for other stand-alone purposes

Dongarra et al. : "High-performance computing: clusters, constellations, MPPs, and future directions", Computing in Science & Engineering (Volume:7 , Issue: 2)

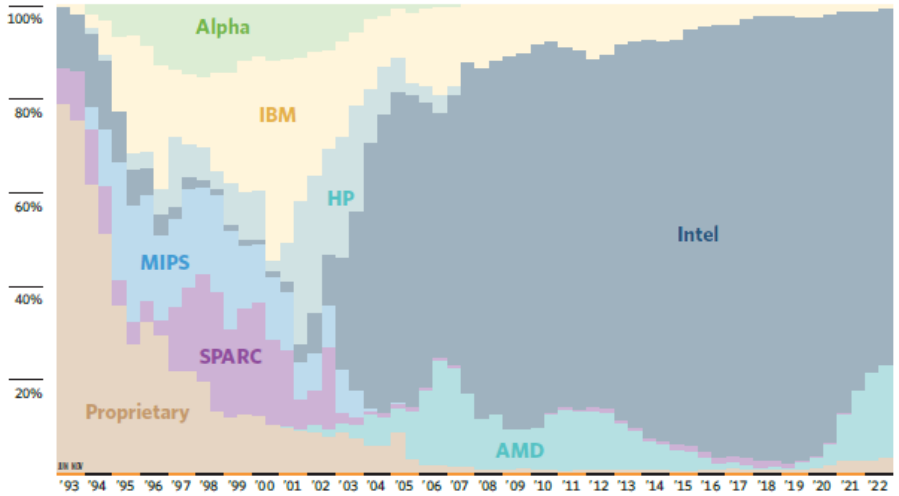
(*) Picture from: http://en.wikipedia.org/wiki/Computer_cluster

Top500.org – stats

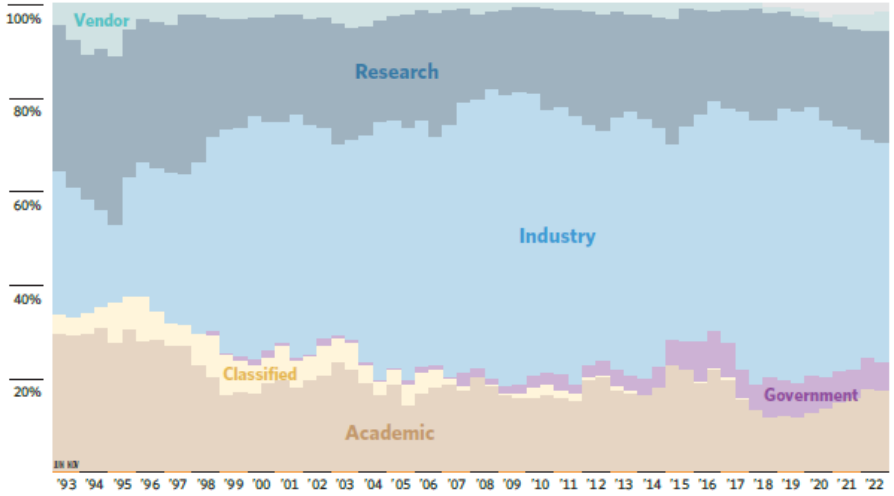
ARCHITECTURES



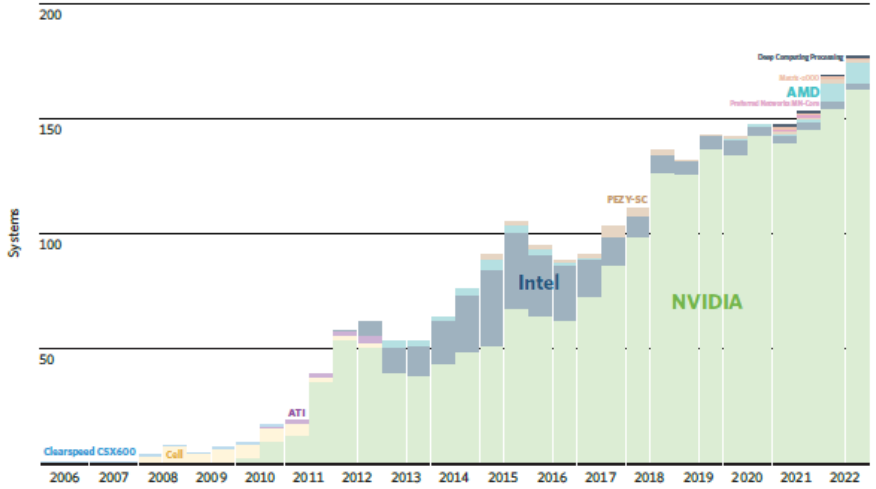
CHIP TECHNOLOGY



INSTALLATION TYPE



ACCELERATORS/CO-PROCESSORS



HPC made easy: **HPC Bubbles**

terabit

- Cluster HPC istanziati su Cloud tramite interfacce user-friendly
- Acquisizione delle bubbles di datacloud tramite progetti PNRR

Generated with AI · May 2024

Gara "HPC Bubbles"

- Accordo Quadro Nazionale
  - Listino prezzi per nodi + accessori
  - Fondi Terabit, DARE e ICSC
  - 2 anni di validità
    - Indizione: 26/05/23
    - Aggiudicazione: 15/02/24
    - Contratto: 26/04/24
  - Lotto1 (base di gara: € 8.680.000,00 + IVA)
    - CPU, GPU, FPGA
    - Ordini
      - 2 Terabit (Nord/Sud): 4,850,685.90 € + IVA
      - 2 DARE (CNAF, BA) : 1,298,276.60 € + IVA
      - 11 ICSC: 2,818,056.60 € + IVA
  - Lotto2 (base di gara: € 2.459.000,00 + IVA)
    - Storage
    - Ordini
      - 2 Terabit (Nord/Sud): 1,344,708.30 €
      - 2 DARE (CNAF ,BA): 513,807.00 € + IVA
      - 5 ICSC: 1,003,105.00 € + IVA

Quantità nodi con fondi Terabit-ICSC-DARE

	Nodo CPU	Nodo GPU	Nodo FPGA Xilinx	Nodo FPGA Terasic	Nodo storage
BA	24	6	0	0	32
CNAF	26	30	2	2	52
MIB	0	0	2	2	0
NA	18	1	2	0	8
PD	6	6	0	0	0
PI	20	0	0	0	0
RM1	12	0	0	0	0
TO	14	6	0	0	0
LNGS	0	6	0	0	12
CT	12	0	0	0	8
LNF	12	0	0	0	0
LNFESA	8	6	0	0	6
LNL	4	0	0	0	0
MI	4	0	0	0	0
TOTALE	160	61	6	4	118

Core: 30 kcore fisici
Circa 34 HS/core

GPU: 244 NVIDIA H100
34 TFLOPS FP64 ➔ 8.3PFLOPS FP64
40 FPGA
InfiniBAnd 400Gbs

45 PB RAW

HPC Bubbles

 Nodo CPU Lenovo Lenovo ThinkSystem SR665 V3	192 core fisici - Dual AMD AMD EPYC 9654 96C 360W 2.4GHz 1.5TB RAM DDR5 IB NDR 400G - NVIDIA ConnectX-7 NDR OSFP400 1-Port PCIe Gen5 x16 InfiniBand Adapter 20TBL (SSD) + dischi di sistema
 Nodo GPU Lenovo ThinkSystem SR675 V3	Come CPU + 4x NVIDIA H100 SXM5 con 80GB HBM3 (non HBM2e come da offerta)
 Nodo FPGA Lenovo ThinkSystem SR675 V3	32core - AMD EPYC 9124 16C 200W 3.0GHz Processor RAM 768GB DDR5 IB NDR 440G 4 x XILINX U55C o 4 x TerasicP0701
 Nodo Storage (CEPH Bricks) DELL PowerEdge R760xd2	64 core fisici – Dual Intel Xeon Gold di quarta generazione (Sapphire Rapids), modello 6428N, frequenza 1.8GHz, 32Core/64Thread, 60M Cache, DDR5-4800, 185W TDP 1TB RAM DDR5 IB Mellanox 400G 384 TBL HDD + 25.6 TBL NVMe
 Accessori	Switch IB, Switch ETH – NVIDIA Modello SN3420 - 12x QSFP28 100GbE + 48x SFP28 25GbE Cavi IB, Cavi ETH Transceiver vari Assistenza 3+2

Quantità nodi HPC BUBBLES con fondi DARE – Terabit per Spoke8 in zona Certifcata ISO27001

	Nodo CPU	Nodo GPU	Nodo FPGA Xilinx	Nodo FPGA Terasic	Nodo storage
BA_DARE	12	6	0	0	6
BA_TerabitS8	0	0	0	0	0
CNAF_DARE	10	9	0	0	16
CNAF_TerabitS8	0-8	0-8	0	0	0-6