

Sept. 23rd, 2025

Managing a Large-Scale HPC Infrastructure: **User eXperience**



Daniele Di Bari ([d.dibari@cineca.it](mailto:d.dibari@ Cineca.it))

Alessandro Marani ([a.marani@cineca.it](mailto:a.marani@ Cineca.it))

User Support & Production Team
High Performance Computing Dept.



PRESENTATION

OUTLINE

- **Presentation of CINECA**
- **How to Request HPC Resources**
- **Leonardo's architecture**
- **Accessing the system (2FA)**
- **Login nodes and accounting**
- **Storage Areas and Data Transfer**

Not for profit **CONSORTIUM**

Since 1969 Cineca supports the Italian Academic System



120 MEMBERS

2 Ministries, 71 Universities,
47 Academic and Research Institutions



6 OFFICES

Bologna, Milan, Rome, Naples, Chieti, Palermo



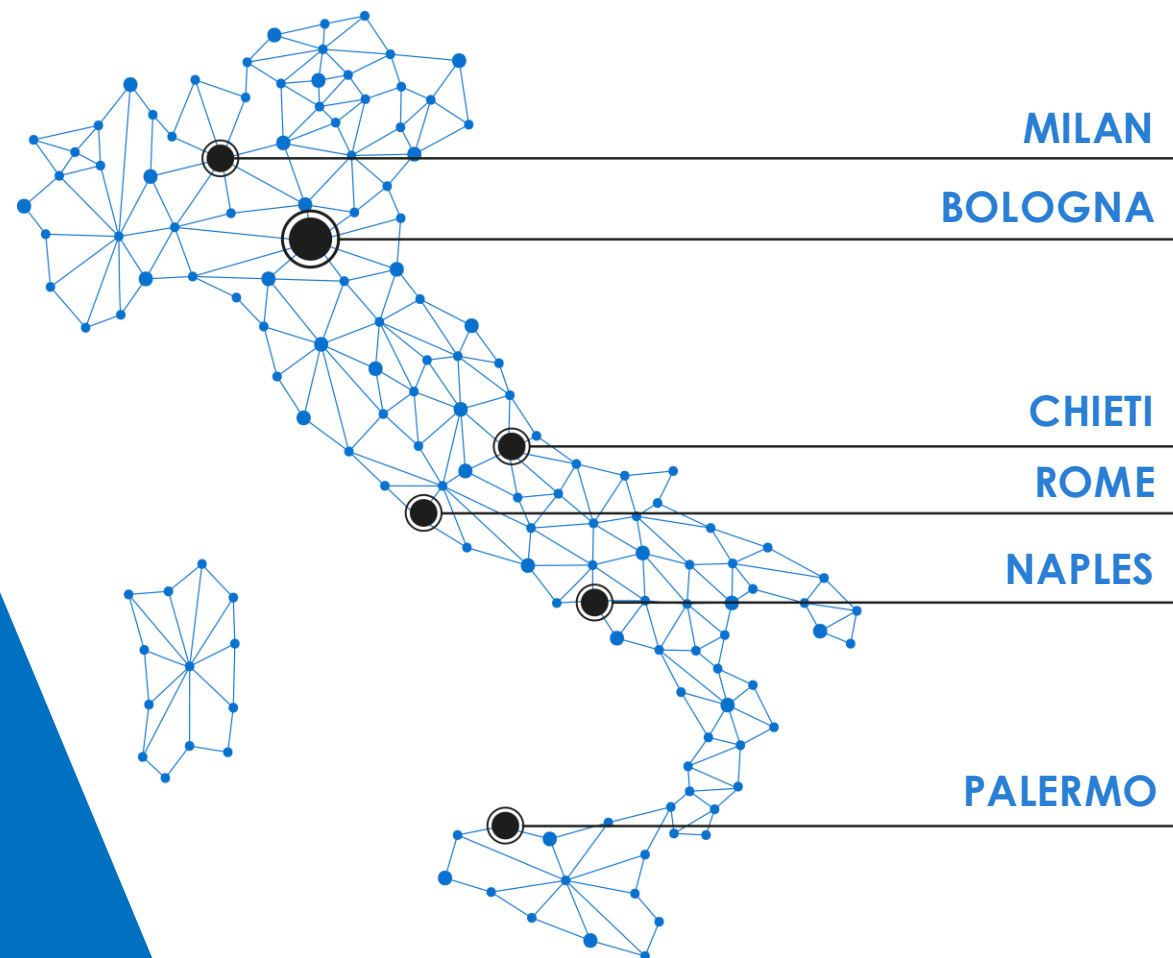
≈200

Employees (in HPC)



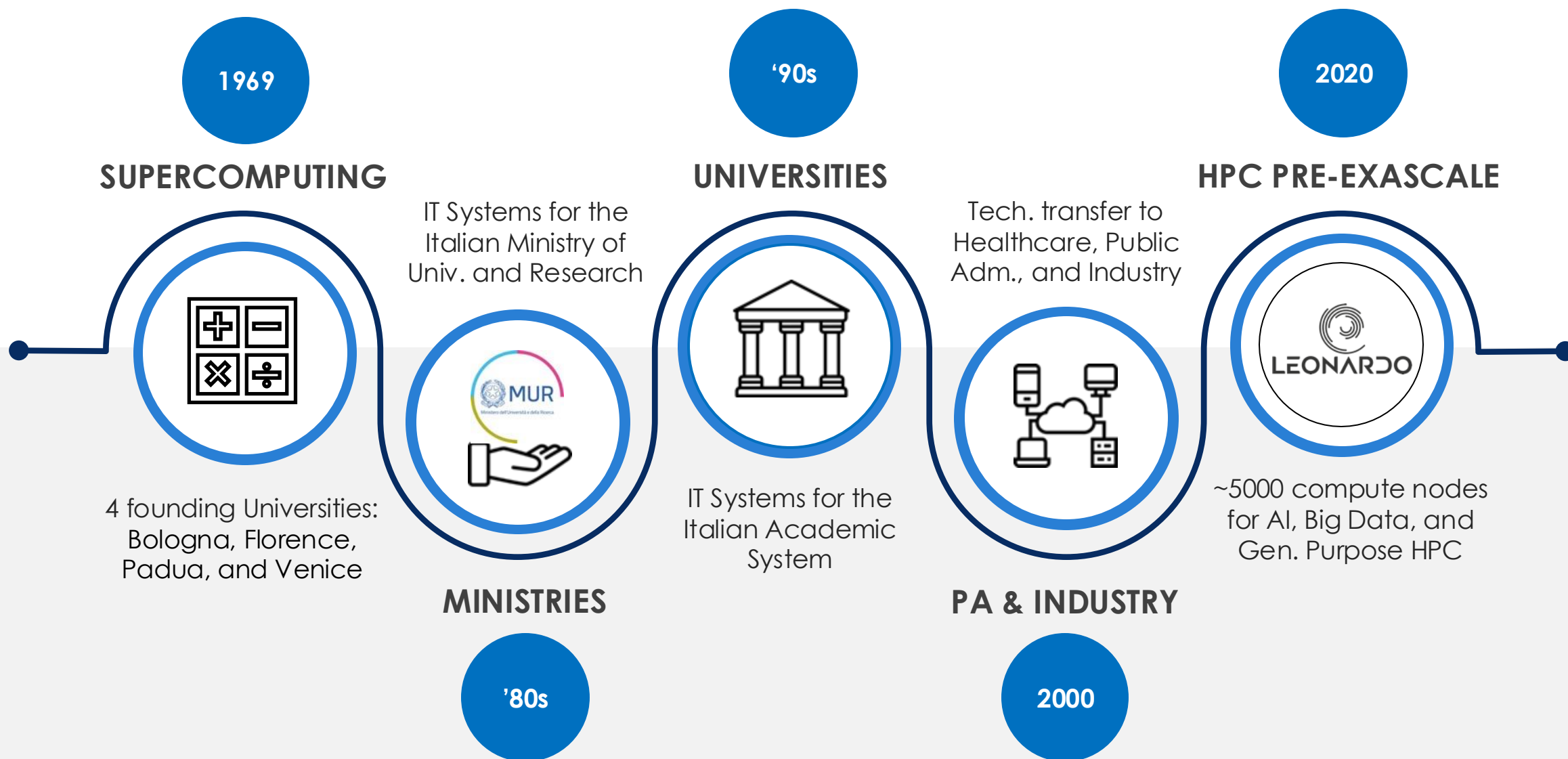
≈ 130 MLN €

Yearly Revenue



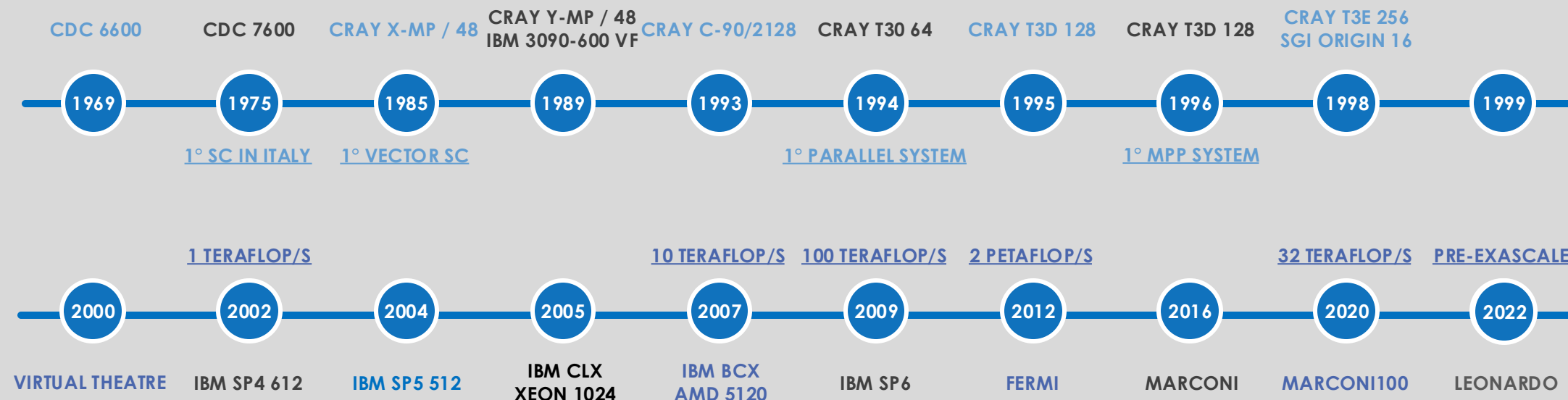
MORE THAN 50 YEARS OF EXPERIENCE

History and Core Competencies of CINECA



MORE THAN 50 YEARS OF EXPERIENCE

Timeline of CINECA HPC Systems



EuroHPC Joint Undertaking

Overview

The EuroHPC JU, established in 2018, coordinates Europe's supercomputing strategy.

Its goals include broadening access to HPC for public and private users and fostering key skills for European science and industry.

| NAME | LOCATION | SUSTAINED PERFORMANCE |
|----------------------|--------------------------------|-----------------------|
| JUPITER | Jülich, <i>GERMANY</i> | 793 |
| LUMI | Kajaani, <i>FINLAND</i> | 386 |
| Leonardo | Bologna, <i>ITALY</i> | 249 |
| MareNostrum 5 | Barcelona, <i>SPAIN</i> | 215 |
| MeluXina | Bissen, <i>LUXEMBOURG</i> | 12.8 |
| Karolina | Ostrava, <i>CZECH REPUBLIC</i> | 9.6 |
| Discoverer | Sofia, <i>BULGARIA</i> | 7.5 |
| Vega | Maribor, <i>SLOVENIA</i> | 6.9 |
| Deucalion | Guimarães, <i>PORTUGAL</i> | 4.5 |

EuroHPC-JU HPC SYSTEMS



Exascale



Pre-exascale



Petascale / mid-range

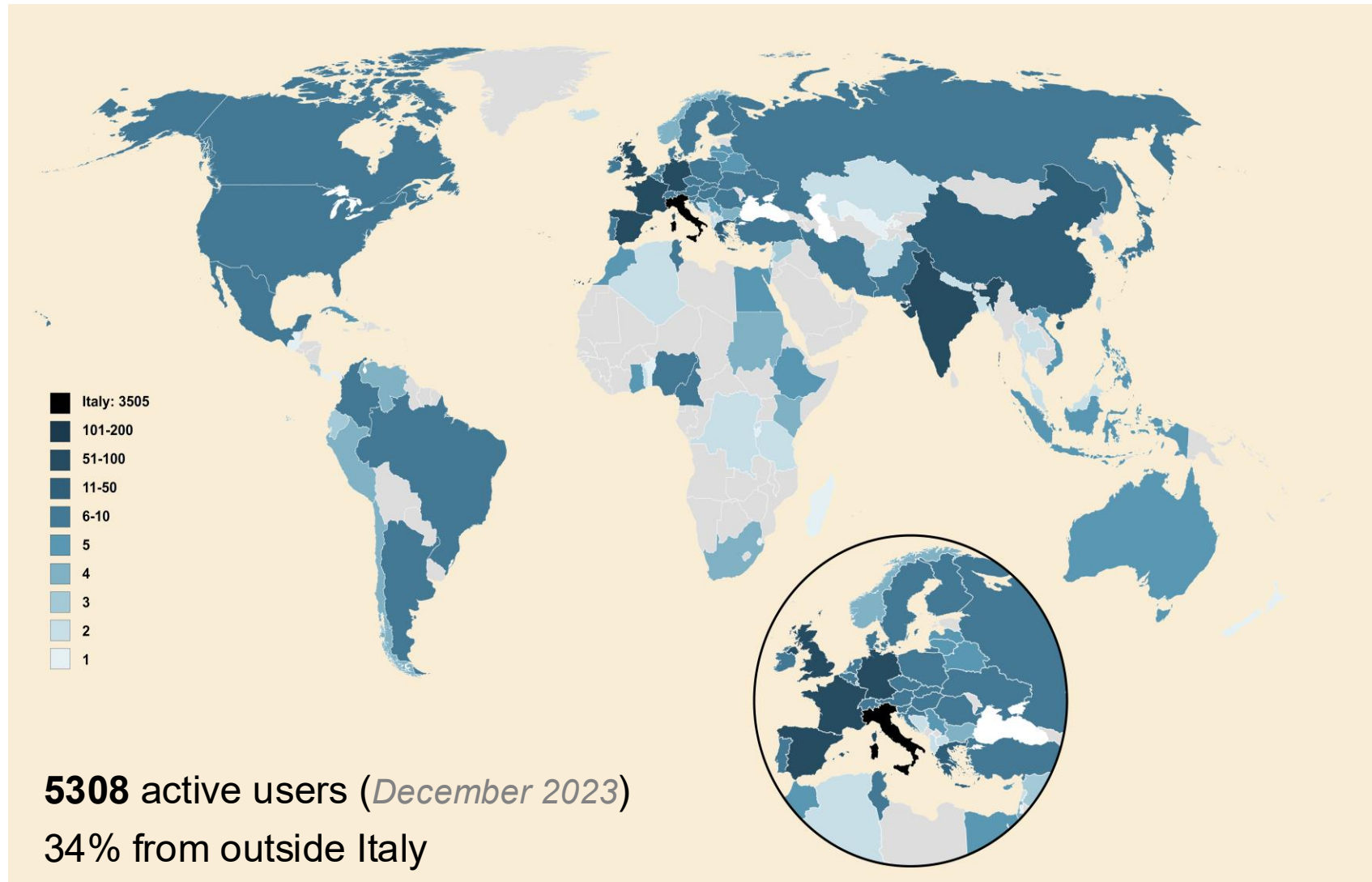


Quantum system



HPC USERS

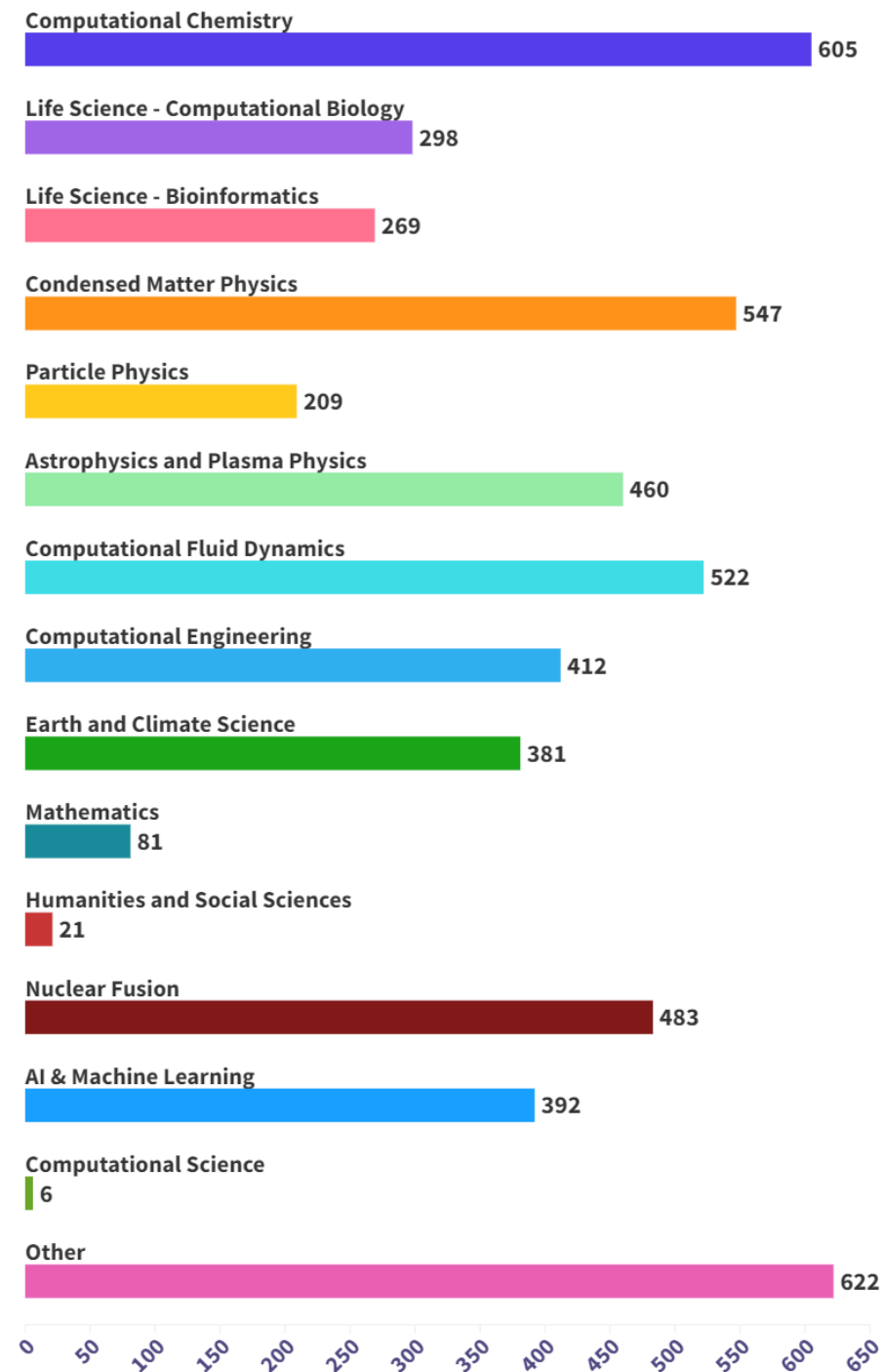
Spatial Distribution



SCIENTIFIC DOMAINS

Number of HPC projects by scientific domains in **2024**.

Areas of interest range from computational chemistry to particle physics, from astrophysics to climate disciplines.



INTERNATIONAL COLLABORATIONS

Scientific Projects, CoE, and more

HPC infrastructure



Digital infrastructure



HPC Center of Excellence



Environment



Big Data / AI / ML



Life Science



Multimedia / Cultural Heritage



Energy efficiency / HPC Technologies Co-Design





PRESENTATION

OUTLINE

- **Presentation of CINECA**
- **How to Request HPC Resources**
- **Leonardo's architecture**
- **Accessing the system (2FA)**
- **Login nodes and accounting**
- **Storage Areas and Data Transfer**



HPC RESOURCE ALLOCATION

CINECA aims and basic principles

OUR OBJECTIVES

- Providing Italian and European researchers with an advanced computational environment.
- Supporting researcher for increasing their competitiveness.
- Soliciting large-scale and computationally intensive projects.

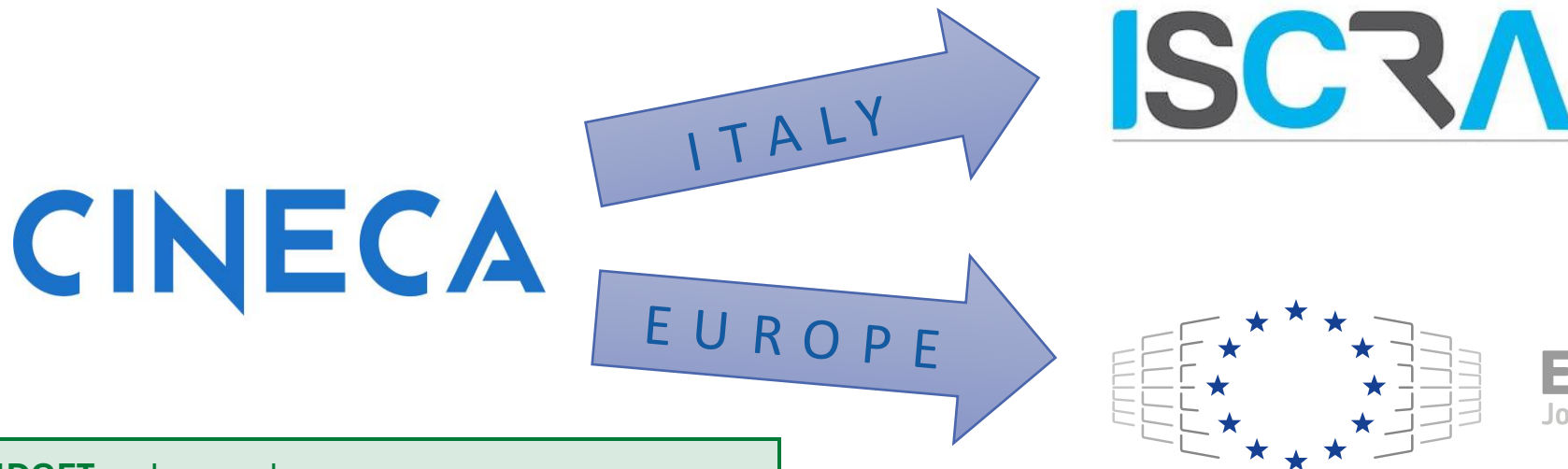
BASIC PRINCIPLES

- Transparency
- Fairness
- Conflict of Interest management
- Confidentiality

HPC RESOURCE ALLOCATION

OVERVIEW

- **Ways to Get Resources:** mainly via peer-reviewed calls (national or international), with exceptions for commercial agreements.
- **Application Process:** Submission of a proposal detailing the scientific use case, resources needed, time usage, and codes to be run. Proposals are reviewed for scientific quality and technical feasibility.



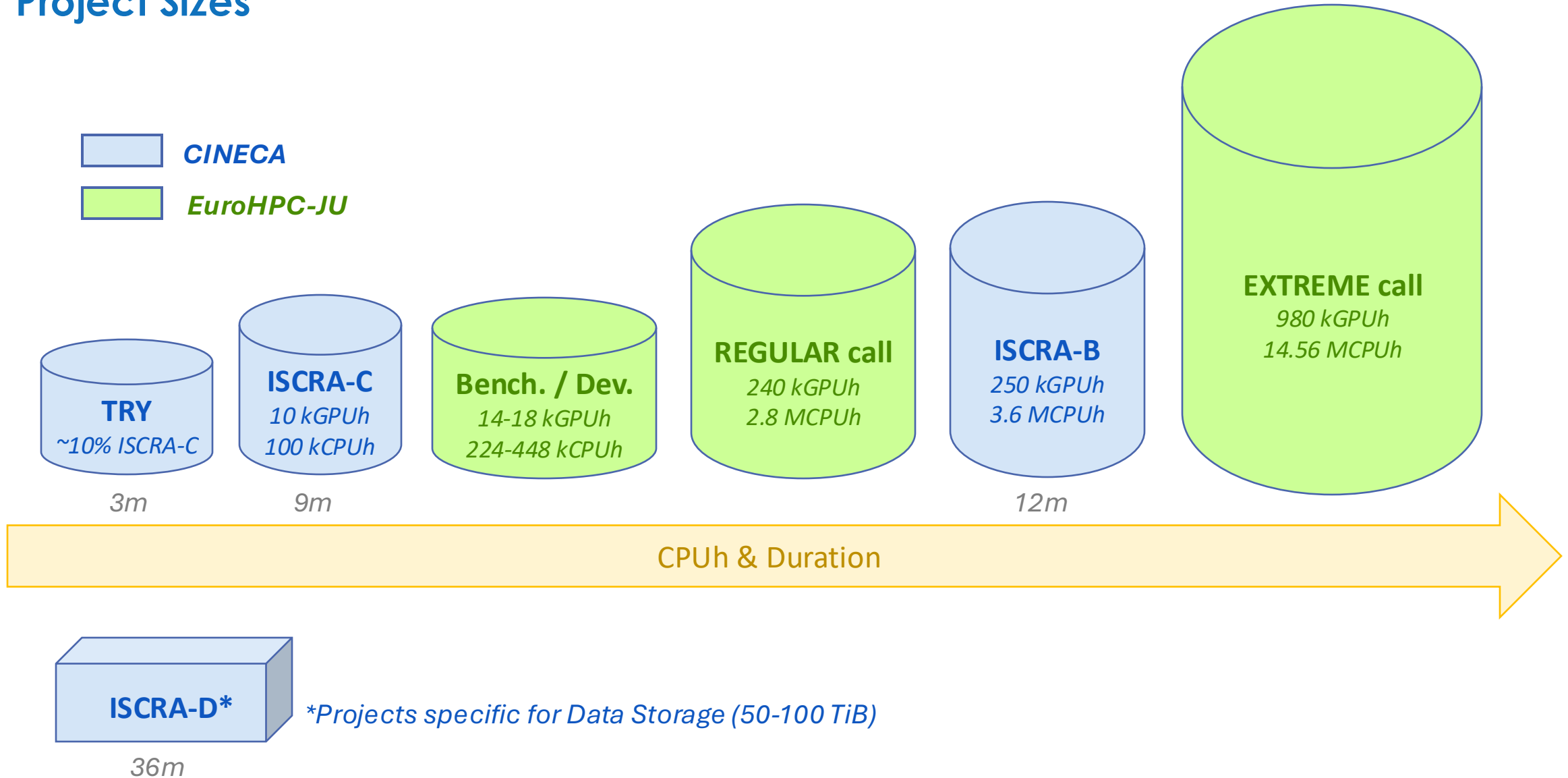
TOTAL BUDGET on Leonardo:

- 1280/1510 MCPUh/year (*non-accelerated partition*)
- 103/121 MGPUh/year (*accelerated partition*)

HPC RESOURCE ALLOCATION

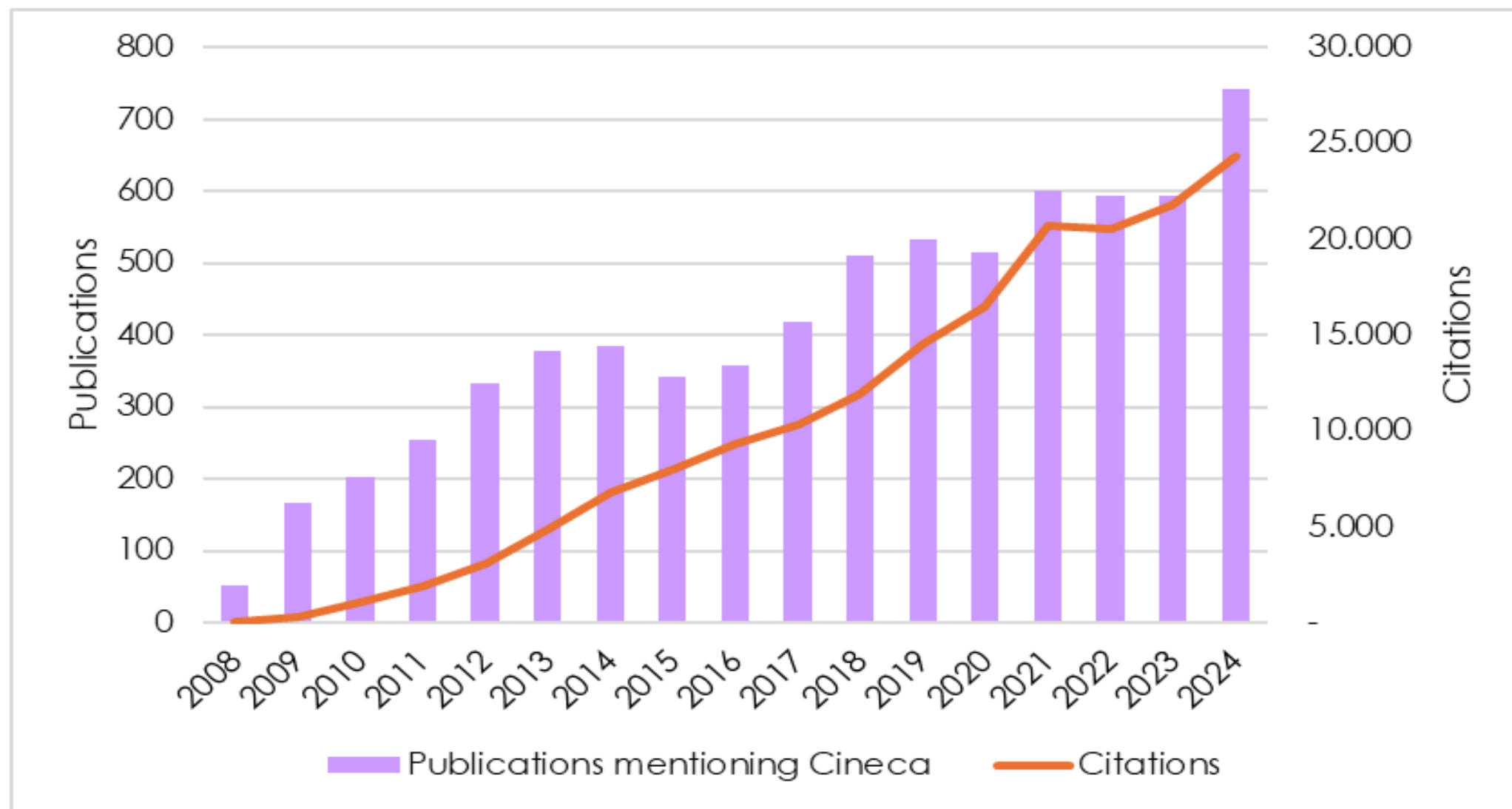
Project Sizes

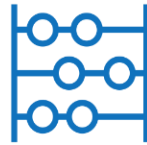
 **CINECA**
 **EuroHPC-JU**



SCIENTIFIC IMPACT

ISCRA



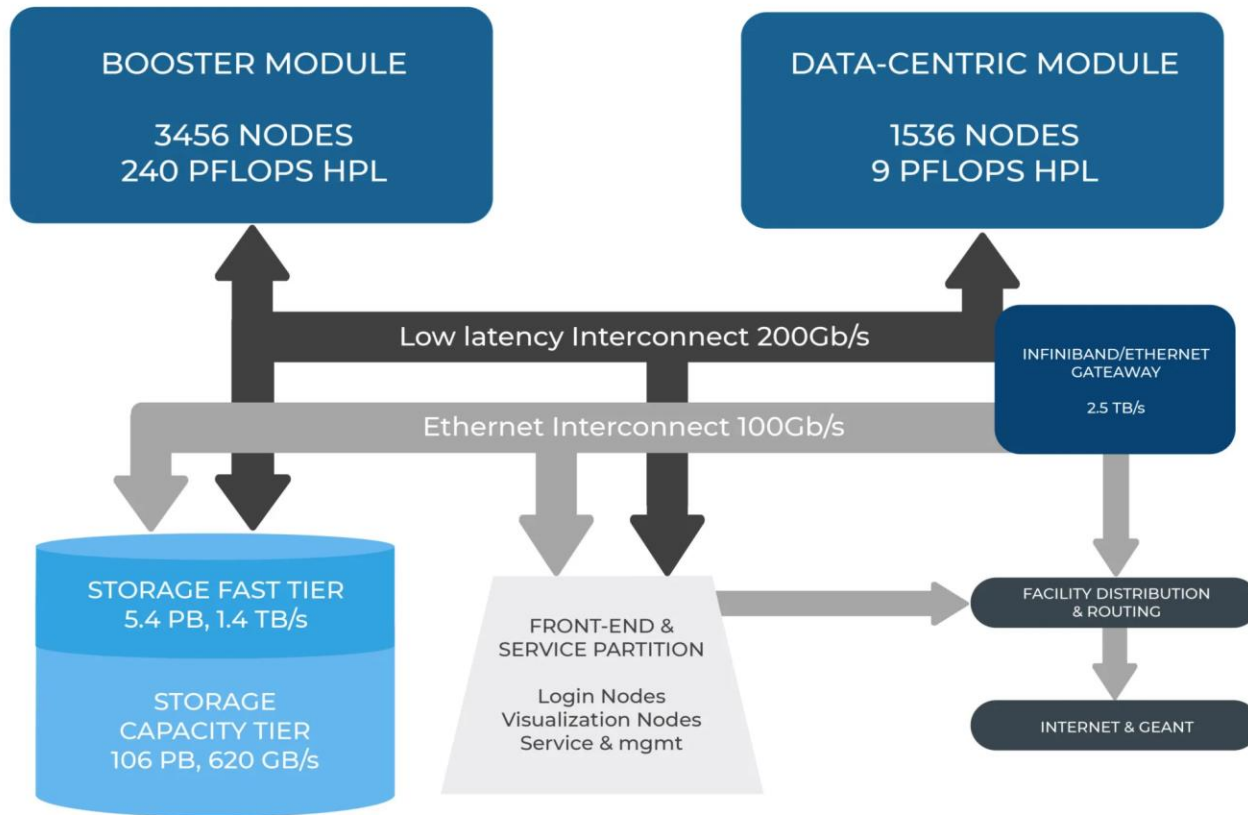


Leonardo's Architecture

Leonardo Supercomputer

LEONARDO's ARCHITECTURE

SYSTEM OVERVIEW & LOGIN NODES



LOGIN NODES:

Processors:

2x Intel Xeon Platinum 8358 Processor
(Intel Ice Lake – 32 cores, 3.4 GHz with Turbo)

- Hyper Threading (×2) is enabled
- RAM: 512 GiB RAM DDR4 3200MHz
- 14TiB disk in RAID1 configuration
- NO GPUs
- Open to outside network
- *Serial* partition on two login node

BOOSTER MODULE (GPU-Accelerated)

Key Features & Specifications



Atos BullSequana X2135 "Da Vinci" blade

- 3456 nodes
- 1 × [Intel Xeon Platinum 8358 Processor](#) (32 cores)
- RAM: 512 (8 x 64) GB DDR4 3200 MHz
- Accelerators: 4 × [GPU NVIDIA A100](#) custom - **15% performance improvement over the standard A100**
- Internal network: NVIDIA Mellanox HDR DragonFly+ 200Gb/s
- **DISKLESS!!!**
- Shared storage (*InfiniBand-Connected*): 106 PiB Capacity tier storage + 5.4 PiB Fast tier storage



Rmax per node: ~90 TFLOPS

Rmax: ~241 PFLOPS

**IN PRODUCTION
SINCE AUGUST 2023**

**Rmax = Maximal LINPACK performance achieved (TOP500)*

BOOSTER MODULE (GPU-Accelerated)

Processor Details

Intel Xeon Platinum 8358 Processor

- **32 cores**, each with 1.25 MiB of L2 cache and 48+32 KiB of L1 cache.
- 2 threads per core if hyperthreading is enable (only on login nodes).
- 2 of AVX-512 FMA Units.
- **48 MiB of L3 cache**, shared across all cores.
- 503 GiB of available RAM, divided into **2 NUMA nodes**.
- Processor Base Frequency: 2.60 GHz
- Max Turbo Frequency: 3.40 GHz

More information is available using the following commands:

```
$ lstopo-no-graphics  
$ numactl -H
```



The **RAM available for user jobs is 494,000 MiB** (*slightly over 482 GiB*), as approximately 20 GiB is reserved for the operating system.

BOOSTER MODULE (GPU-Accelerated)

Processor Details

```
[<username>@lrdnXXXX ~]$ lstopo-no-graphics
Machine (503GB total) + Package L#0 + L3 L#0 (48MB)
```

Proc. Units

```
Group0 L#0
```

```
NUMANode L#0 (P#0 251GB)
```

```
L2 L#0 (1280KB) + L1d L#0 (48KB) + L1i L#0 (32KB) + Core L#0 + PU L#0 (P#0)
```

```
L2 L#1 (1280KB) + L1d L#1 (48KB) + L1i L#1 (32KB) + Core L#1 + PU L#1 (P#1)
```

```
[...]
```

```
L2 L#30 (1280KB) + L1d L#30 (48KB) + L1i L#30 (32KB) + Core L#30 + PU L#30 (P#30)
```

```
L2 L#31 (1280KB) + L1d L#31 (48KB) + L1i L#31 (32KB) + Core L#31 + PU L#31 (P#31)
```

```
[<username>@lrdnXXXX ~]$ numactl -H
```

NUMA nodes

```
available: 2 nodes (0-1)
```

```
node 0 cpus: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
```

```
node 0 size: 256926 MB
```

```
node 0 free: 237337 MB
```

```
node 1 cpus: 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
```

```
node 1 size: 257999 MB
```

```
node 1 free: 242985 MB
```

```
node distances:
```

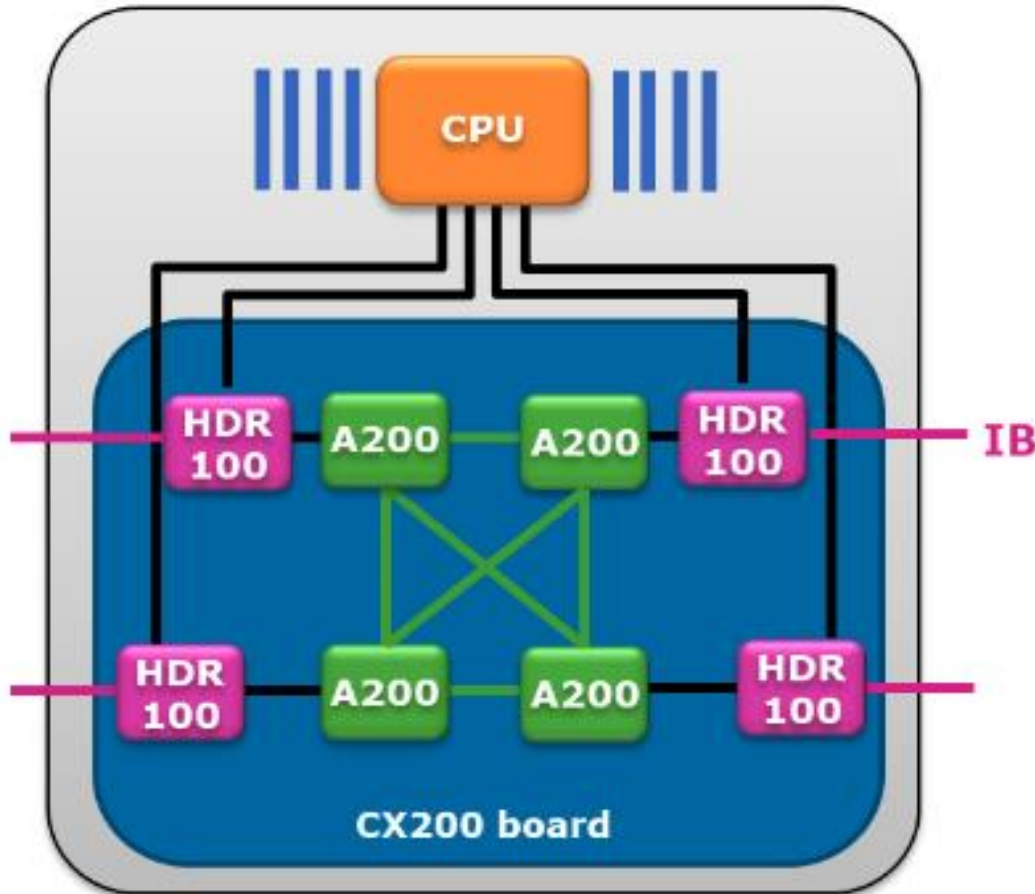
```
node 0 1
```

```
0: 10 11
```

```
1: 11 10
```

BOOSTER MODULE (GPU-Accelerated)

GPU Interconnect & Memory Architecture



GPU peak performance

- 11.2 TFlops Peak FP64 per GPU or....
- 22.4 TFlops Peak FP32 per GPU

Intra-node connections

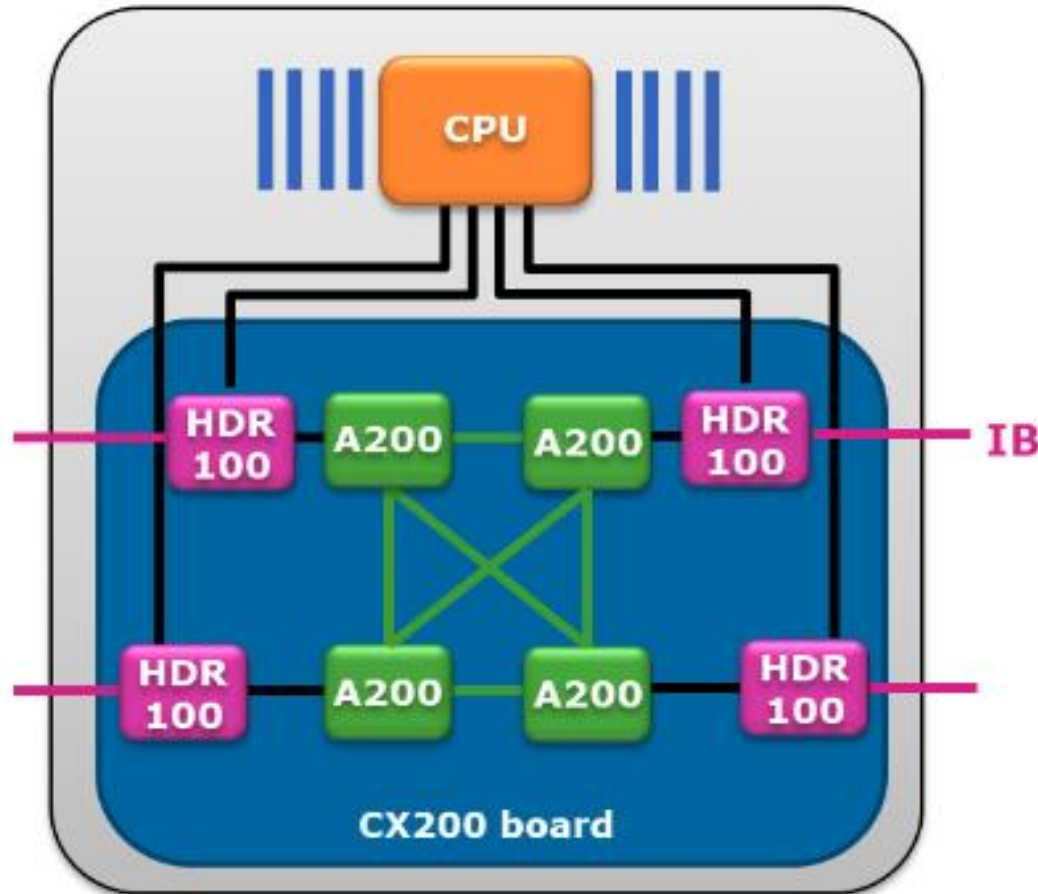
- NVLink, PCIe, GPU direct
- 200 GB/s between the GPU pairs
- Each GPU has direct 100Gb/s connection to the InfiniBand network
- PCIe Gen4 @ 31.5 GB/s

Memory

- 6.5 TB/s GPU memory bandwidth

BOOSTER MODULE (GPU-Accelerated)

GPU Interconnect & Memory Architecture



How to select the NIC:

```
$ cat select_nic_ucx
#!/bin/bash
export UCX_NET_DEVICES=mlx5_$SLURM_LOCALID:1
exec $*
$ srun ./select_nic_ucx my_application
```

For more information use the command "ucx_info -c", or see the [UCX documentation](#).



Single-socket design, so no concerns about GPU-to-CPU binding. Equal access from each core to the GPU for balanced performance.

BOOSTER MODULE (GPU-Accelerated)

GPU Interconnect & Memory Architecture

- 4x NVLink 3.0 per GPU pair, **200 GB/s bi-directional bandwidth**.
- **256 GB HBM2e total**, >6.5 TB/s memory bandwidth across 4 GPUs.
- **No NVLink between CPU-GPU** (Intel CPUs don't support NVLink); PCIe 4.0 used instead.
- Each GPU connects to a Mellanox HDR100 ConnectX-6 NIC via PCIe passthrough, enabling full-speed CPU-to-GPU communication and GPU-direct communications over InfiniBand.
- ConnectX-6 provides 32 PCIe Gen4 lanes (16 lanes to CPU, 16 lanes to GPU).



Summary

- Direct CPU & GPU connection to the network for lowest latency.
- Four HDR links for high-speed cluster interconnect.
- More power-efficient node design by eliminating external PCIe switches.

DCGP MODULE (CPU)

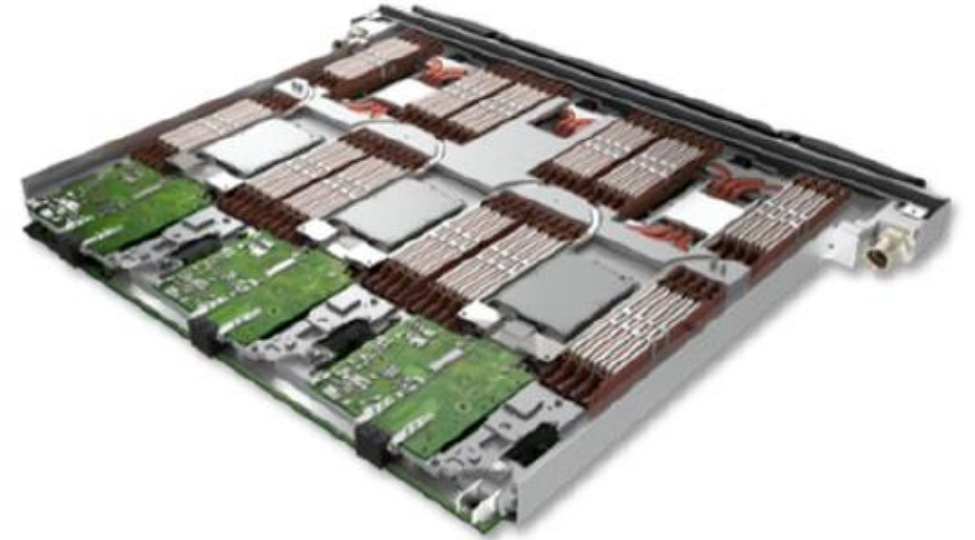
Key Features & Specifications

BullSequana X2140 three-node CPU Blade

- 1536 nodes (512 blades)
- 2 × [Intel Xeon Platinum 8480+ Processor](#) (56 cores each)
- RAM: 512 GB DDR5 4800 MHz
- Infiniband: 1 × NVIDIA HDR cards 100 Gbps via PCIe Gen 5
- Disk: 1 M.2 SSD 3.84 TB



Rmax per node: ~8.5 TFLOPS
Rmax: ~13 PFLOPS



**IN PRODUCTION
SINCE FEBRUARY 2024**

**Rmax = Maximal LINPACK performance achieved (TOP500)*

DCGP MODULE (CPU)

Processor Details

Compute Node:

- **112 cores** in total (56 cores per socket).
- 503 GiB of available RAM, divided into **8 NUMA nodes** (4 per socket).

Intel Xeon Platinum 8480+ Processor:

- **56 cores**, each with 2 MiB of L2 cache and 48+32 KiB of L1 cache.
- 2 of AVX-512 FMA Units.
- **105 MiB of L3 cache**, shared across all cores.
- Processor Base Frequency: 2.00 GHz
- Max Turbo Frequency: 3.80 GHz

More information is available using the following commands:

```
$ lstopo-no-graphics  
$ numactl -H
```



The **RAM available for user jobs is 494,000 MiB** (*slightly over 482 GiB*), as approximately 20 GiB is reserved for the operating system.

DCGP MODULE (CPU)

Processor Details

```
[<username>@lrdnXXXX ~]$ lstopo-no-graphics
```

```
Machine (503GB total) + Package L#0 + L3 L#0 (105MB)
```

```
Group0 L#0
```

```
NUMANode L#0 (P#0 62GB)
```

```
L2 L#0 (2048KB) + L1d L#0 (48KB) + L1i L#0 (32KB) + Core L#0 + PU L#0 (P#0)
```

```
L2 L#1 (2048KB) + L1d L#1 (48KB) + L1i L#1 (32KB) + Core L#1 + PU L#1 (P#1)
```

```
L2 L#2 (2048KB) + L1d L#2 (48KB) + L1i L#2 (32KB) + Core L#2 + PU L#2 (P#2)
```

```
L2 L#3 (2048KB) + L1d L#3 (48KB) + L1i L#3 (32KB) + Core L#3 + PU L#3 (P#3)
```

```
L2 L#4 (2048KB) + L1d L#4 (48KB) + L1i L#4 (32KB) + Core L#4 + PU L#4 (P#4)
```

```
L2 L#5 (2048KB) + L1d L#5 (48KB) + L1i L#5 (32KB) + Core L#5 + PU L#5 (P#5)
```

```
[...]
```

```
L2 L#106 (2048KB) + L1d L#106 (48KB) + L1i L#106 (32KB) + Core L#106 + PU L#106 (P#106)
```

```
L2 L#107 (2048KB) + L1d L#107 (48KB) + L1i L#107 (32KB) + Core L#107 + PU L#107 (P#107)
```

```
L2 L#108 (2048KB) + L1d L#108 (48KB) + L1i L#108 (32KB) + Core L#108 + PU L#108 (P#108)
```

```
L2 L#109 (2048KB) + L1d L#109 (48KB) + L1i L#109 (32KB) + Core L#109 + PU L#109 (P#109)
```

```
L2 L#110 (2048KB) + L1d L#110 (48KB) + L1i L#110 (32KB) + Core L#110 + PU L#110 (P#110)
```

```
L2 L#111 (2048KB) + L1d L#111 (48KB) + L1i L#111 (32KB) + Core L#111 + PU L#111 (P#111)
```

Proc. Units

DCGP MODULE (CPU)

Processor Details

```
[<username>@lrdnXXXX ~]$ numactl -H
available: 8 nodes (0-7)
node 0 cpus: 0 1 2 3 4 5 6 7 8 9 10 11 12 13
node 0 size: 63569 MB
node 0 free: 1417 MB
[...]
node 7 cpus: 98 99 100 101 102 103 104 105 106 107 108 109 110 111
node 7 size: 64505 MB
node 7 free: 55128 MB
node distances:
node  0  1  2  3  4  5  6  7
 0:  10 12 12 12 21 21 21 21
 1: 12 10 12 12 21 21 21 21
 2: 12 12 10 12 21 21 21 21
 3: 12 12 12 10 21 21 21 21
 4: 21 21 21 21 10 12 12 12
 5: 21 21 21 21 12 10 12 12
 6: 21 21 21 21 12 12 10 12
 7: 21 21 21 21 12 12 12 10
```

NUMA nodes

STORAGE

Key Features & Specifications

Fast Tier

5.4 PB @ 1.4 TB/s

SSD disks (NVMe)
(home + fast scratch)



Capacity Tier

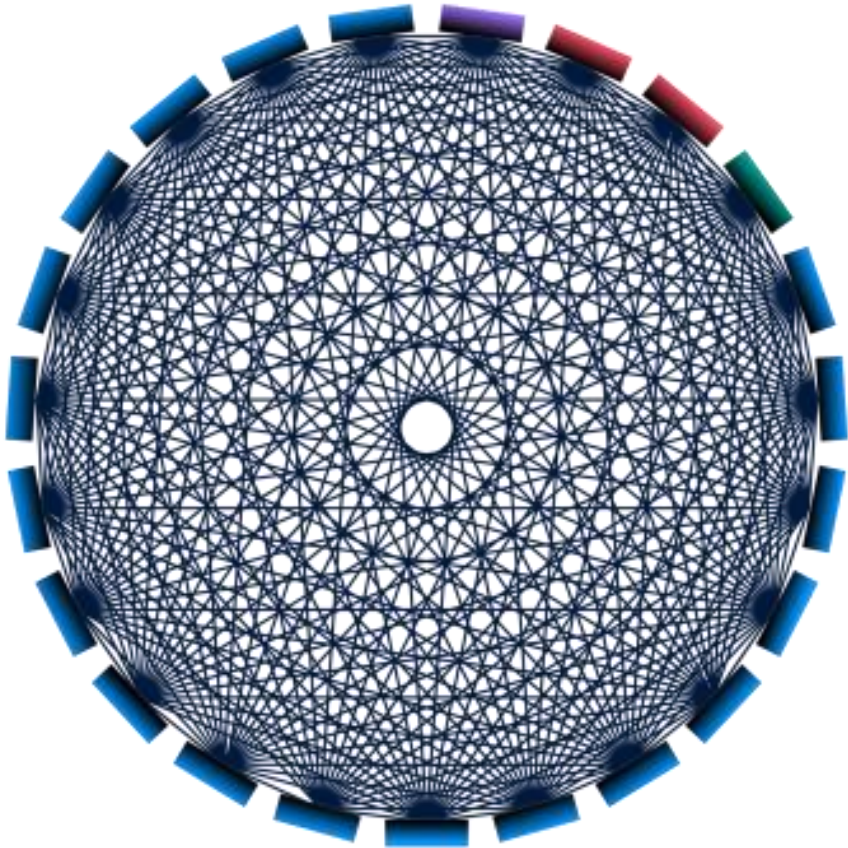
106 PB @ read 744 GB/s - write 620 GB/s

HDD disks
(work + large scratch + DRES)



INTER-NODE NETWORK TOPOLOGY

Key Features & Specifications



Booster Cells

DCGP Cells

Hybrid Cell

Booster + DCGP

Service Cell

Dragonfly+ topology

Based on NVidia Mellanox InfiniBand High Data Rate (HDR) and [NVIDIA Quantum QM8700 switches](#)

- All nodes are divided into cells.
- Cells are connected in an all-to-all topology with 18 independent connections between two different cells.
- Within each cell, a non-blocking two-layer fat tree topology is employed.

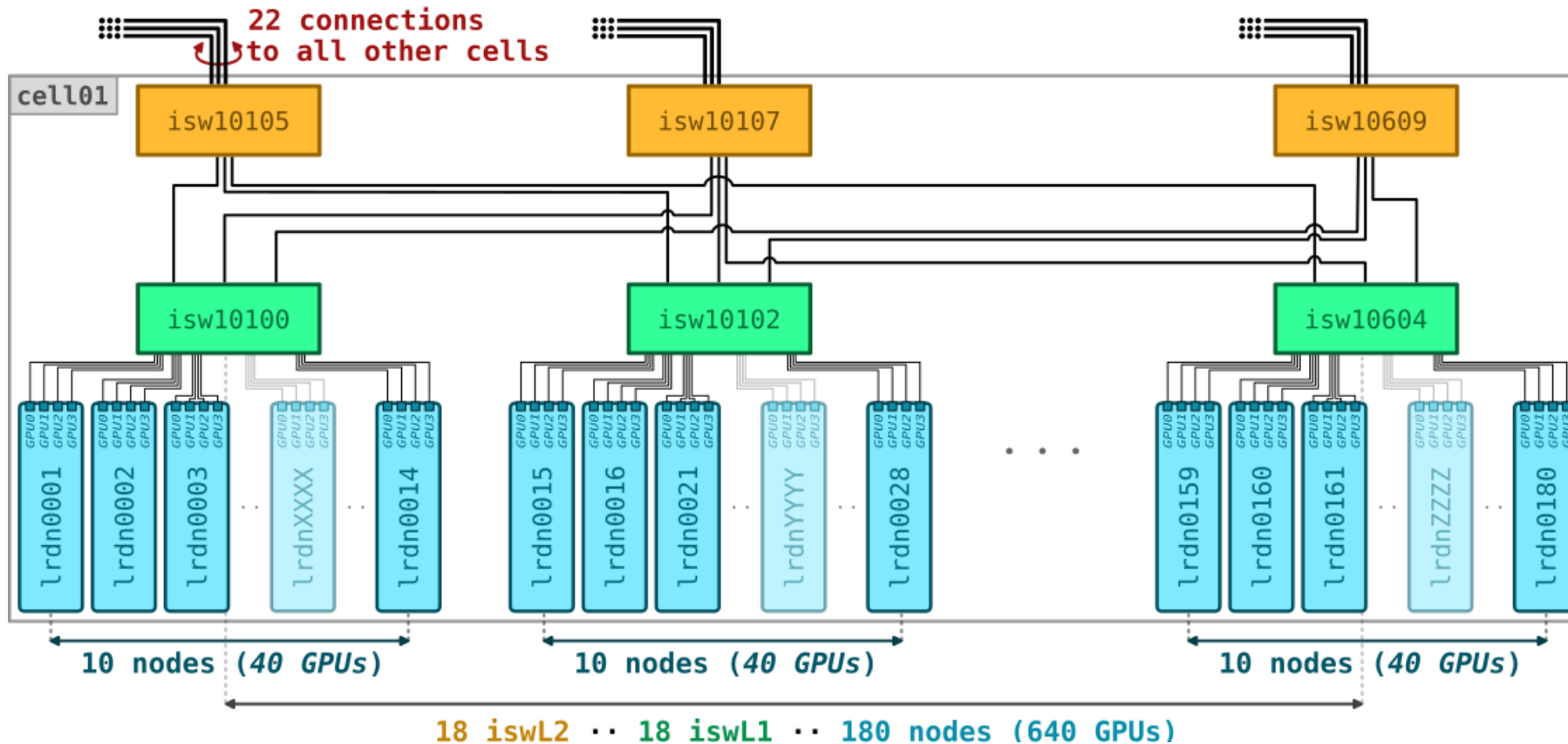
Slurm Optimization & Network Adaptability

- Slurm is aware of the topology and tune the node allocations on the dragonfly+ network.
- Adaptive Routing Algorithm enabled to alleviate network congestion.

INTER-NODE NETWORK TOPOLOGY

Booster Cell

All-to-All (btw. cells) / Two-Level Fat Tree (inside the cells)



LEVEL 2 SWITCHES (iswL2):

downlinks = $18 \times 200\text{Gb/s}$

uplinks = $22 \times 200\text{Gb/s}$

LEVEL 1 SWITCHES (iswL1):

downlinks = $40 \times 100\text{Gb/s}$

uplinks = $18 \times 200\text{Gb/s}$

COMPUTE NODES:

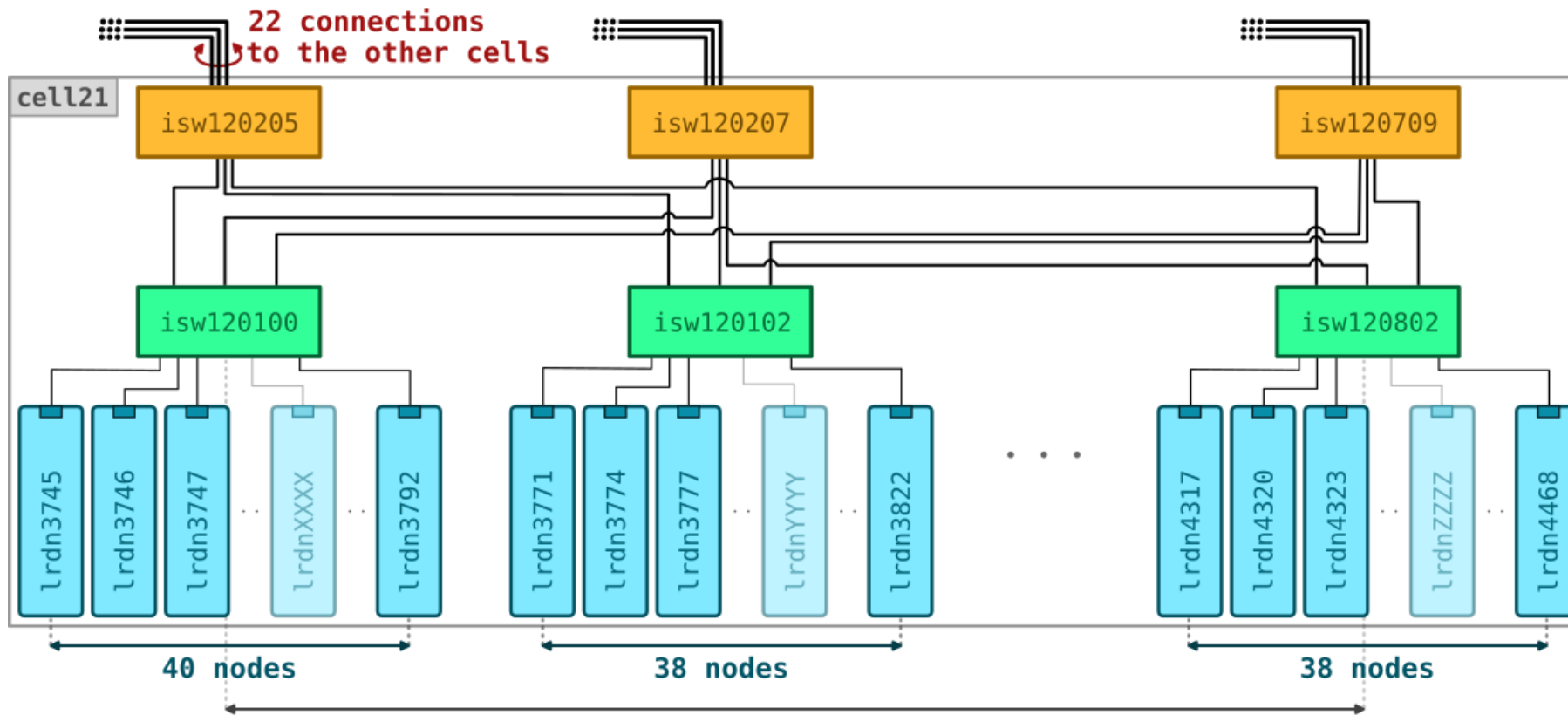
$2 \times 200\text{Gb/s}$ dual port NICs

(one 100Gb/s port per GPU)

INTER-NODE NETWORK TOPOLOGY

DCGP Cell

All-to-All (btw. cells) / Two-Level Fat Tree (inside the cells)



LEVEL 2 SWITCHES (iswL2):

downlinks = $16 \times 200\text{Gb/s}$

uplinks = $22 \times 200\text{Gb/s}$

LEVEL 1 SWITCHES (iswL1):

downlinks $\sim 40 \times 100\text{Gb/s}$

uplinks = $18 \times 200\text{Gb/s}$

COMPUTE NODES:

$1 \times 100\text{Gb/s}$ NICs

18 iswL2 .. 16 iswL1 .. 624 nodes

INTER-NODE NETWORK TOPOLOGY

Service Cell & Summary

- Cell dedicated to I/O servers and front-end nodes.
- 13 x L1 switches [*18 uplinks 200Gbps, only Fast Tier has 200Gbps links other nodes of this cell have 100Gbps links*].
- 18 x L2 switches [*18 downlinks 200gbps + 22 uplinks 200Gbps*].



Summary

- L2 and L1 switches support two configuration modes: 40-port mode at 200Gbps, and 80-port mode at 100Gbps.
- Mixed port configuration is allowed – a single 200Gbps port can be split into 2 × 100Gbps ports.

TOP500 SUPERCOMPUTERS

www.top500.org

June 2025



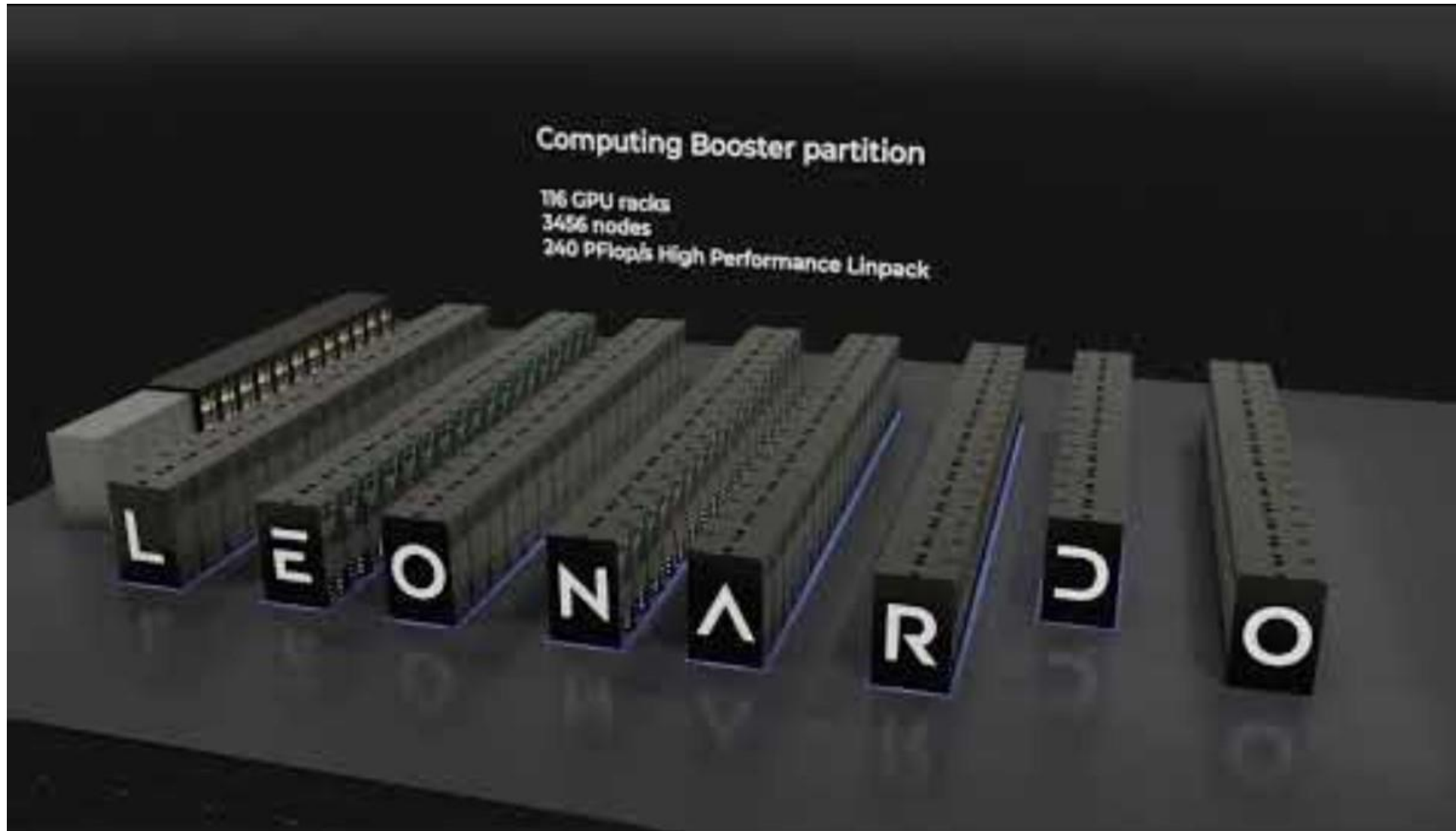
| # | SYSTEM | COUNTRY | Rmax* [PFLOPS] | POWER [kW] |
|----|-----------------|-------------|-------------------|---------------|
| 1 | El Capitan | USA | 1,742.00 | 29,581 |
| 2 | Frontier | USA | 1,353.00 | 24,607 |
| 3 | Aurora | USA | 1,012.00 | 38,698 |
| 4 | Jupiter Booster | Germany | 793.40 | 13,088 |
| 5 | Eagle | USA | 561.20 | -- |
| 6 | HPC6 | Italy | 477.90 | 8,461 |
| 7 | Fugaku | Japan | 442.01 | 29,899 |
| 8 | Alps | Switzerland | 434.90 | 7,124 |
| 9 | LUMI | Finland | 379.70 | 7,107 |
| 10 | Leonardo | Italy | 241.20 | 7,494 |

On November 2022 Leonardo was in the 4th place in the top 500 classification for the first time, been after Frontier, Fugaku and LUMI

*Rmax = Maximal LINPACK performance achieved

LEONARDO VIRTUAL TOUR

System Overview



LEONARDO VIRTUAL TOUR

Cooling System





PRESENTATION

OUTLINE

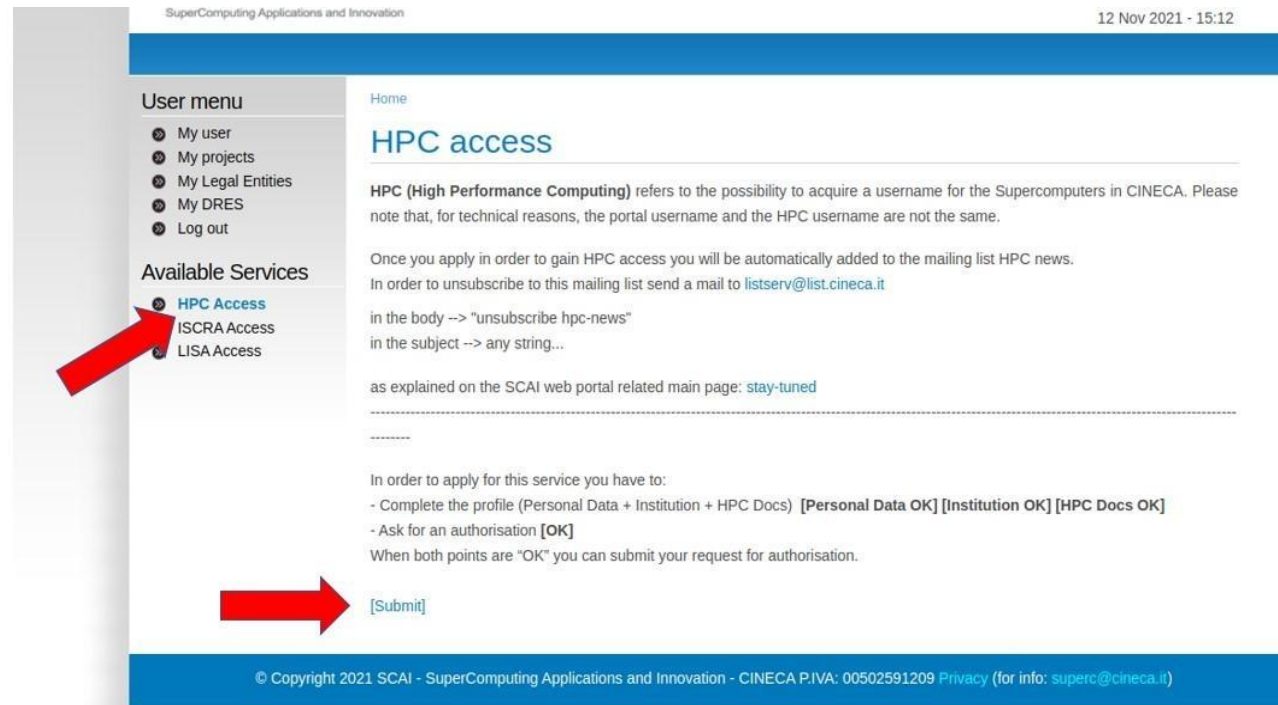
- Presentation of CINECA
- How to Request HPC Resources
- Leonardo's architecture
- **Accessing the system (2FA)**
- **Login nodes and accounting**
- **Storage Areas and Data Transfer**

HPC access

✓ If you are a new HPC user:

If you are a new collaborator of an **existing** HPC project, register on userdb.hpc.cineca.it and ask the PI of the project to add you as collaborator

Follow the steps in the **"HPC access"** subsection of your UserDB account page (provide **information** about **yourself** and your **institution**, upload a valid **ID document**), and finally click "[submit]" to request HPC access



The screenshot shows the 'HPC access' page within a web portal. The page has a blue header with the text 'SuperComputing Applications and Innovation' and a timestamp '12 Nov 2021 - 15:12'. On the left, there is a 'User menu' with options: 'My user', 'My projects', 'My Legal Entities', 'My DRES', and 'Log out'. Below this is an 'Available Services' section with three items: 'HPC Access' (highlighted with a red arrow), 'ISCRA Access', and 'LISA Access'. The main content area is titled 'HPC access' and contains the following text: 'HPC (High Performance Computing) refers to the possibility to acquire a username for the Supercomputers in CINECA. Please note that, for technical reasons, the portal username and the HPC username are not the same. Once you apply in order to gain HPC access you will be automatically added to the mailing list HPC news. In order to unsubscribe to this mailing list send a mail to listserv@list.cineca.it in the body --> "unsubscribe hpc-news" in the subject --> any string... as explained on the SCAI web portal related main page: [stay-tuned](#)'. Below this, it states: 'In order to apply for this service you have to: - Complete the profile (Personal Data + Institution + HPC Docs) [Personal Data OK] [Institution OK] [HPC Docs OK] - Ask for an authorisation [OK] When both points are "OK" you can submit your request for authorisation.' At the bottom of the main content area, there is a '[Submit]' button, which is pointed to by a red arrow. The footer of the page contains the copyright information: '© Copyright 2021 SCAI - SuperComputing Applications and Innovation - CINECA P.IVA: 00502591209 Privacy (for info: superc@cineca.it)'.

Identity Provider: registration

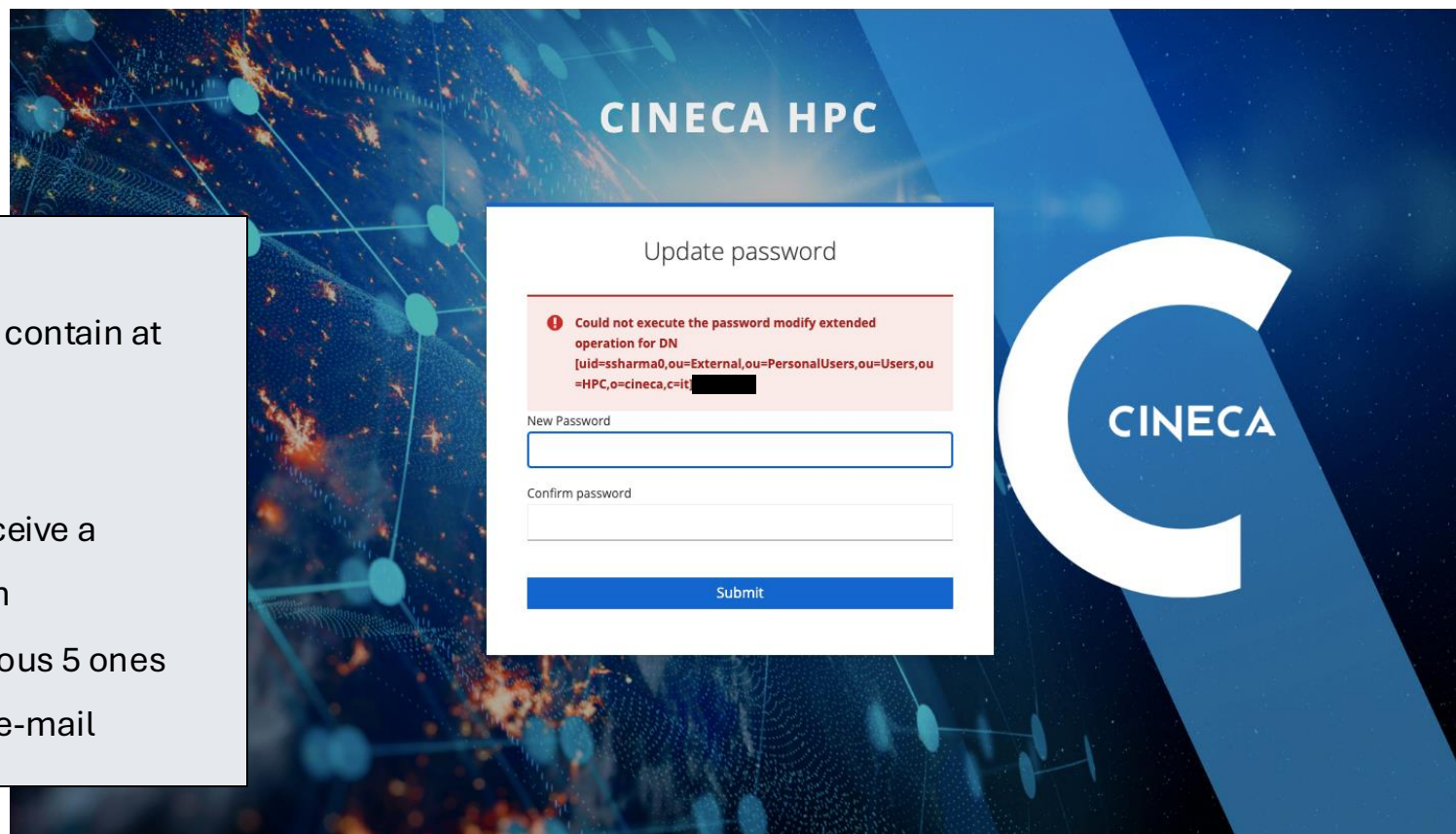
✓ IP setup: HPC password creation/update

You will need to create or update your HPC password respecting the **policies** for password definition.

If you try to set a password that does not respect the policies you will get a somewhat cryptic error, shown below.

Policies for password definition:

- The new password has to be 10 characters long and contain at least 1 capital letter, 1 number, and 1 special character (!"#\$%&'()*+,-./:;<=>?@[\\]^_`{|}~)
- The password has a validity of 3 months. You will receive a reminder 10 days before the expiration when you login
- The new password has to be different from the previous 5 ones
- Any password change will be notified to the user by e-mail

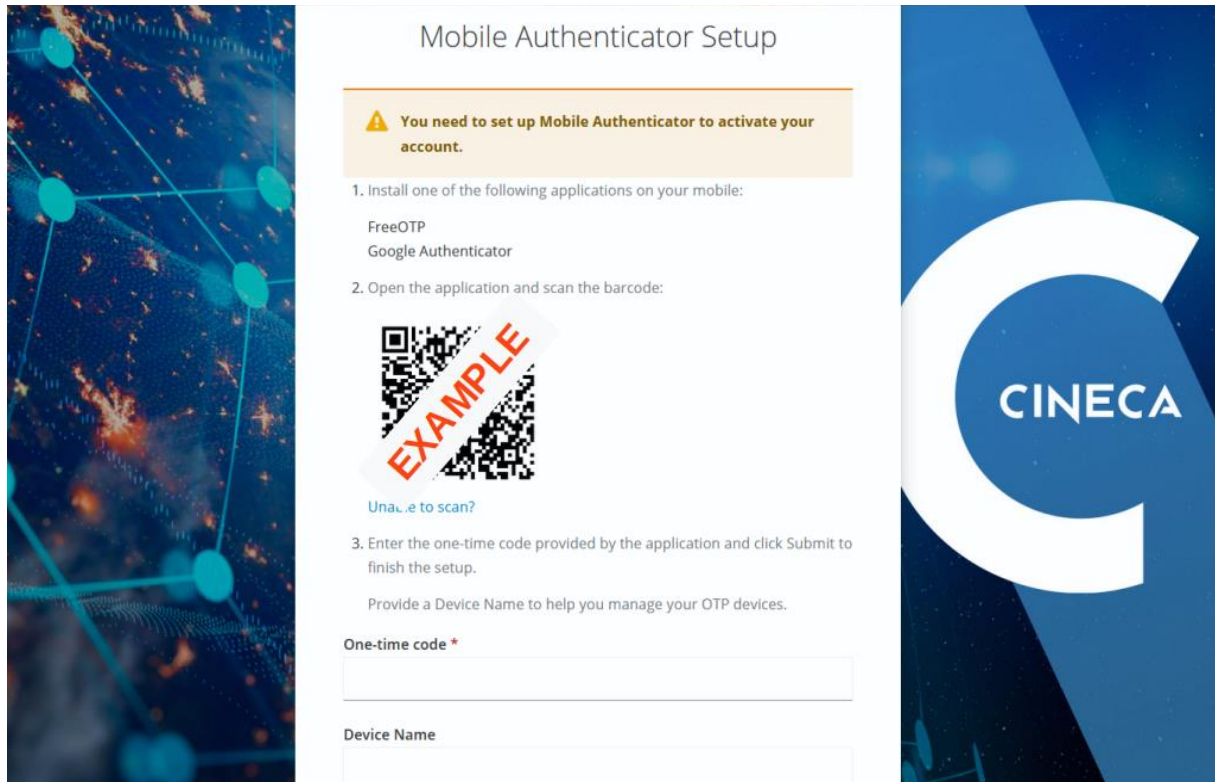


Identity Provider: registration

✓ IP setup: configuring the OTP generator

After that, you will have to setup an OTP generator app for your account.

It is **strongly** advised to install the app on a second device, like a smartphone.



The image shows a 'Mobile Authenticator Setup' form on the left and the CINECA logo on the right. The form has a blue header with the title 'Mobile Authenticator Setup'. Below the header is a yellow warning box with a triangle icon and the text 'You need to set up Mobile Authenticator to activate your account.' The form contains three numbered steps: 1. Install one of the following applications on your mobile: FreeOTP, Google Authenticator. 2. Open the application and scan the barcode: A QR code is shown with a red 'EXAMPLE' watermark. 3. Enter the one-time code provided by the application and click Submit to finish the setup. Below the steps are two input fields: 'One-time code *' and 'Device Name'. The CINECA logo is a large white 'C' on a blue background with the word 'CINECA' in white text.

OTP generator configuration steps:

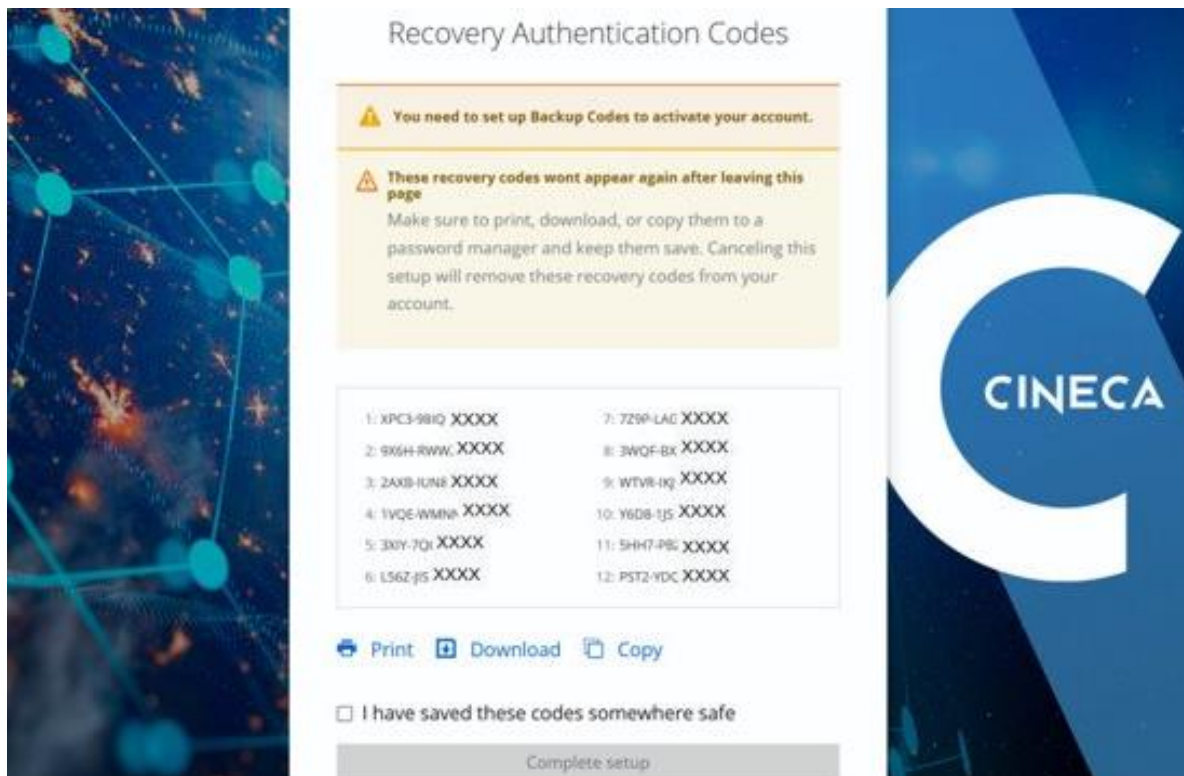
- Install OTP generator app on second device (Google Authenticator, Microsoft Authenticator, FreeOTP, ...)
- Scan the QR code from the app or click on "Unable to scan" and insert the displayed code on the app
- Insert the OTP code displayed in the app on the Mobile Authenticator Setup form

Identity Provider: registration

✓ IP setup: configuring the OTP generator

After the configuration is complete the page will display **12 recovery codes**.

Make sure to **save** these codes somewhere (**ALL** of them) by downloading, printing or copying them in a text file.



The screenshot shows the 'Recovery Authentication Codes' page. It features a warning message: 'You need to set up Backup Codes to activate your account.' and another warning: 'These recovery codes wont appear again after leaving this page'. Below these warnings, there is a table of 12 recovery codes, each consisting of a number followed by a 16-character alphanumeric string. At the bottom, there are buttons for 'Print', 'Download', and 'Copy', a checkbox for 'I have saved these codes somewhere safe', and a 'Complete setup' button.

| Recovery Authentication Codes | |
|-------------------------------|--------------------|
| 1: XPC3-98IQ XXXXX | 7: 7Z9P-LAG XXXXX |
| 2: 9X6H-RWW XXXXX | 8: 3WQF-BX XXXXX |
| 3: 2AXB-IJNR XXXXX | 9: WTVR-IQ XXXXX |
| 4: 1VQE-WMNA XXXXX | 10: Y6DB-IJS XXXXX |
| 5: 3XY-7QI XXXXX | 11: 5HH7-PBC XXXXX |
| 6: L56Z-JS XXXXX | 12: PST2-YDC XXXXX |

☐ I have saved these codes somewhere safe

Complete setup

At this point the Identity Provider configuration is **complete**.

You will be able to manage **e-mail address** (synced with UserDB), **HPC password**, **OTP generators** and **OTP recovery codes** from the account management section on sso.hpc.cineca.it.

Also, you will be able to **reset** your password from the Identity Provider login page.

Identity Provider: FAQs and notes



Problem: the OTP code is rejected as not valid.



Solution: make sure that the time of your device is properly synchronized.

Google Authenticator has a "**Time correction for codes**" function in its settings useful for this purpose.

P.S.: make sure to input the OTP code **without** hyphens ("-") and spaces!!



If your e-mail address begins with a digit, we'll have to change it.

Currently there is a **bug** in the authentication software, and e-mail addresses that start with numerical characters do not work.

So, for the time being, if your e-mail address begins with a number, we will ask you an **alternative address**, and change it for you on UserDB.

Smallstep configuration: FAQs and notes



DO NOT launch the CA bootstrap command with admin privileges

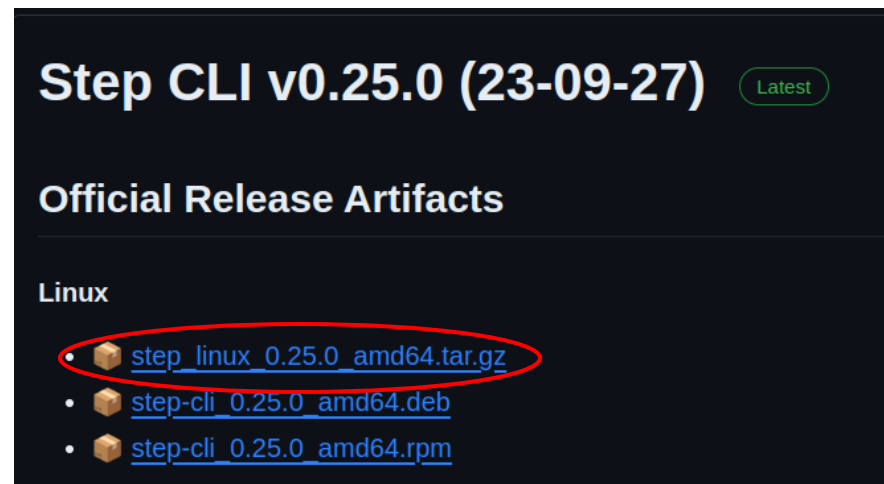
```
$ step ca bootstrap --ca-url=https://sshproxy.hpc.cineca.it --  
fingerprint 2ae1543202304d3f434bdc1a2c92eff2cd2b02110206ef06317e70c1c1735ecd
```

It needs to be launched with the **same** permissions as the unprivileged user.



Do I need admin privileges to install smallstep?

No, you can download a compressed archive from the GitHub repo releases page and **unpack it in your user space**. Then you can simply add its **./bin** folder to your system's **PATH**.



Smallstep: usage



Get the temporary ssh certificate in the ssh-agent:

Running the command:

```
$ step ssh login <user_mail> --provisioner cineca-hpc
```

Will open your browser and prompt you to authenticate with **username**, **password** and **OTP code** on the Identity Provider website.

The ssh-agent service will hold in its memory the **temporary SSH certificate** (valid for **12 hours**) and you can login with:

```
$ ssh <username>@login.leonardo.cineca.it
```

and also authenticate on Leonardo with every program that uses the SSH protocol (**scp**, **rsync**, **VSCode Remote**, ...)

Smallstep: usage



Downloading the certificate to the local PC filesystem:

```
$ step ssh certificate <user_mail> --provisioner cineca-hpc id_ecdsa
```

Will also open your browser and prompt you to authenticate with **username**, **password** and **OTP code** on the Identity Provider website.

This time three files **id_ecdsa**, **id_ecdsa.pub** and **id_ecdsa-cert.pub** will be downloaded to your filesystem (**encrypting** the private key is **strongly recommended**), you can then login either with:

```
$ ssh -i <path_to_id_ecdsa> <username>@login.leonardo.cineca.it
```

Or add the key directly to the ssh-agent and connect with:

```
$ ssh-add <path_to_id_ecdsa>  
$ ssh <username>@login.leonardo.cineca.it
```

Smallstep: usage



Useful commands:

```
$ ssh-add -L  
$ step ssh list
```

Will list all SSH keys in the memory of the SSH agent.

```
$ step ssh list --raw '<user_email>' | step ssh inspect
```

Will show details about the temporary certificate held by the SSH agent, and can be used to check if it is valid.

```
$ ssh-add -D
```



PRESENTATION

OUTLINE

- **Presentation of CINECA**
- **How to Request HPC Resources**
- **Leonardo's architecture**
- **Accessing the system (2FA)**
- **Login nodes and accounting**
- **Storage Areas and Data Transfer**

Login nodes

\$ ssh username@login.leonardo.cineca.it

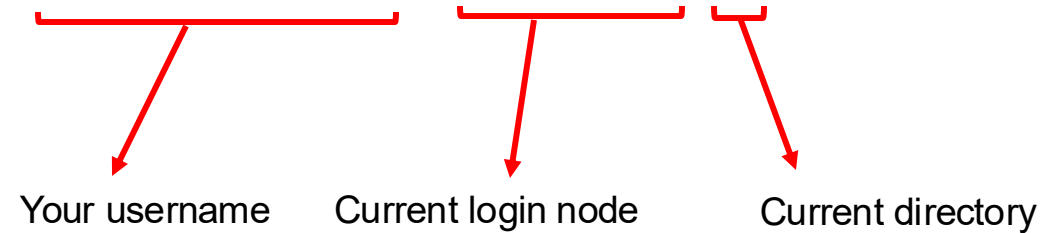
```
mguernel@login02:~
Welcome to:
Leonardo
*****
* Red Hat Enterprise Linux 8.7 (Ootpa)
*
* Booster module:
* Atos Bull Sequana X2135 "Da Vinci" Blade
* 3456 compute nodes with:
*   - 32 cores Ice Lake at 2.60 GHz
*   - 4 x NVIDIA Ampere A100 GPUs, 64GB
*   - 512 GB RAM
*
* DataCentric General Purpose module (DCGP):
* Atos BullSequana X2140 Blade
* 1536 compute nodes with:
*   - 2x56 cores Intel Sapphire Rapids at 2.00 GHz
*   - 512 GB RAM
*
* Internal Network: Nvidia Mellanox HDR DragonFly++
* SLURM 22.05
*
* For a guide on Leonardo:
* https://wiki.u-gov.it/confluence/display/SCAIUS/UG3.2%3A+LEONARDO+UserGuide
* For support: superc@cinca.it
*****
IN EVIDENCE:
- A new personal area $PUBLIC is available to share installations and/or
  data. Please, keep in mind that the $PUBLIC directory is by default open
  to everybody on the cluster, and your files are visible to all users.
- The automatic cleaning of the $SCRATCH area is NOT active at the moment
- RCM will be available soon
- Spack module is available to customize your software environment.
  "module av spack" to list the available versions and
  "module load spack/<version>" to use a specific one
Register this system with Red Hat Insights: insights-client --register
Create an account or view all your systems at https://red.ht/insights-dashboard
Last login: Wed Feb 21 10:38:51 2024 from 84.220.4.202
[mguernel@login02 ~]$
```

Motto of the day

- Short system description
- System status
- “In evidence” messages
- “Important” messages

Bash shell

[<username>@login0X ~]\$



Login nodes

Leonardo (as the other CINECA HPC clusters) is shared among many users, so a

responsible use is crucial!

Login nodes

- The purpose of login nodes is mainly to perform small operations and submitting computing jobs;
- Interactive runs on login nodes are strongly discouraged and should be limited to short test runs:
- ***10 minutes cpu-time limit***
- Avoid running large and parallel applications on login nodes;
- ***No GPUs on login nodes.***

Accounting

To use our clusters you will need HPC **username** associated to an **account**.

Username: Identifies the individual connecting to the system.

Username are **strictly personal**. Every user entitled with login credentials is to be considered personally responsible for any misuse that should take place.

Typically, usernames are:

- 8 characters long
- Composed using the first letter of your name and the first 7 letters of your surname
- Trailing zeros if your surname is less than 7 letters

(e.g. **Mario Rossi** -> **mrossi00**).

Accounting

Account

- Identifies the resource allocation which you can use for your work.
- A **budget** is associated with an account and reports how many resources (computing hours) are available on each cluster.
- An amount of storage is also associated with an account, available on the \$WORK space
- The account budget and storage is shared between all the users that are associated to the account (collaborators).

A single username can use **multiple accounts**, and an account can be used by **multiple usernames**, all competing for the same budget.

Whenever an account **runs out of budget** (in CPU hours), or when its expiring date is met, all the usernames referring to that account won't be able to submit batch jobs anymore.

Accounting: How much will my job cost?

Leonardo Booster

$$\text{Accounted Resources}(\text{cpus-h}) = \text{Reserved Cores eq.}(\text{cpus}) \cdot \text{Elapsed Time}(h)$$

1 Node contains:

- 32 cores (CPUs)
- 4 GPUs
- 494000 MiB RAM

32 CPUs / 4 GPUs

32 CPUs / 494000 MiB RAM



1 GPU = 8 CPUs_{eq} (1/4 Node)



15437.5 MiB RAM = 1 CPU_{eq}

The **accounting** considers:

- Number of allocated **CPUs**
- Amount of allocated **RAM memory**
- Allocated GRES (*Generic Trackable RESources*), such as **GPUs**, TMPFS, exc.

Expressed using the **number of equivalent cores** and takes the **maximum** of these values.

Example 1:

2 hours run with: 1 Node, 16 CPUs, mem=80000, 1 GPU

- CPUs=16
- Mem=80000/15438=5 **Maximum=16**
- GPUs=1*8=8

Budget consumed=2 hours*1 node*16=32 hours

Example 2:

1,5 hours run with: 3 Nodes, 4 CPUs, mem=70000, 4 GPUs

- CPUs=4
- Mem=70000/15438=4.5 **Maximum=32**
- GPUs=4*8=32

Budget consumed=1,5 hours*3 nodes*32=144 hours

Accounting: Monitoring your budget

```
[mguernel@login02 ~]$ saldo -b
```

| account | start | end | total (local h) | localCluster Consumed(local h) | totConsumed (local h) | totConsumed % | monthTotal (local h) | monthConsumed (local h) |
|------------|----------|----------|--------------------|-----------------------------------|--------------------------|------------------|-------------------------|----------------------------|
| cin_staff | 20110323 | 20300323 | 200000002 | 19358765 | 53866669 | 26.9 | 864553 | 492640 |
| cin_propro | 20220427 | 20301231 | 500000 | 2659 | 2786 | 0.6 | 4731 | 0 |
| cin_saldo | 20230524 | 20300323 | 10 | 0 | 0 | 0.0 | 0 | 0 |
| cin_sudo | 20230524 | 20300323 | 10 | 0 | 0 | 0.3 | 0 | 0 |

```
[mguernel@login02 ~]$
```

Account validity period

Total CPU-h on
every cluster

CPU-h consumed
on current cluster

Total CPU-h consumed
on every cluster

Monthly budget
(total and consumed)

To check the buget you have at your disposal you can use the command: **\$ saldo -**

b

Accounting: Budget Linearization

Every account has a **monthly quota** ($\text{total_budget} / \text{total_no_of_months}$).

When users start to consume the account budget, the jobs submitted from the account will **gradually lose priority**, until the monthly budget is fully consumed.

When this happens, you can still run jobs (so it is possible to consume more than the monthly quota each month), but these jobs will have the lowest priority.



PRESENTATION

OUTLINE

- Presentation of CINECA
- How to Request HPC Resources
- Leonardo's architecture
- Accessing the system (2FA)
- Login nodes and accounting
- **Storage Areas and Data Transfer**

Filesystems

\$HOME

- 50 GB per user
- User specific
- Permanent (till user is active)
- Daily backup (soon)

\$PUBLIC

- 50 GB per user
- User specific (permissions **755**)
- Permanent (till user is active)
- **No** backup

\$FAST

Same policies of WORK with faster R/W (SSD)

Data resources (DRES)

Shared area among different projects platforms.

\$WORK

- Quota per account (default 1TB)
- Account specific
- Permanent (account + 6 month)
- **No** backup

\$SCRATCH

- No quota
- User specific
- Temporary (data will be removed after 40 days, and no backup)

Local SSD storage (3TB)

Exploitable in jobs via \$TMPDIR (SLURM directive).
Serial & DCGP only.

All the filesystems are based on **Lustre**

Filesystems

```
[mguernel@login02 ~]$ cindata
USER      AREADESCR      AREAID      FRESH      USED      QTA      USED%      aUSED      aQTA      aUSED%
mguernel  /leonardo_scratch/fast/cin_propro  leonardo_scratch_fast-22057231  35min      --      --      --%      4K      1T      0.0%
mguernel  /leonardo_work/IscrB_SoDi-PSV_0     leonardo_work-20059220          35min      --      --      --%      26T      35T      74.5%
mguernel  /leonardo_scratch/fast/cin_sudo     leonardo_scratch_fast-22058717  35min      --      --      --%      4K      1T      0.0%
mguernel  /leonardo_work/cin_staff            leonardo_work-20042960          35min      --      --      --%      38T      100T     38.1%
mguernel  /leonardo_scratch/fast/cin_saldo    leonardo_scratch_fast-22058716  35min      --      --      --%      4K      1T      0.0%
mguernel  /leonardo_work/cin_propro           leonardo_work-20057231          35min      --      --      --%      4K      1T      0.0%
mguernel  /leonardo_work/cin_sudo             leonardo_work-20058717          35min      --      --      --%      4K      1T      0.0%
mguernel  /leonardo_work/cin_saldo            leonardo_work-20058716          35min      --      --      --%      4K      1T      0.0%
mguernel  /leonardo_scratch/fast/cin_staff    leonardo_scratch_fast-22042960  35min      --      --      --%      5.3G     1T      0.5%
mguernel  /leonardo/home/userinternal/mguernel leonardo_home-10126046          35min      747M     50G     1.5%      --      --      --%
mguernel  /leonardo/pub/userinternal/mguernel leonardo_pub-12126046           34min      10G     50G     21.5%     --      --      --%
mguernel  /leonardo_scratch/large/userinternal/mguernel leonardo_scratch-11126046      33min      121G     --      --%      --      --      --%
```

Area location (full path)

ID

Last update of
cindata

User
occupied
space

User
quota

User
occupied
space
(%)

Shared usage
and quota for the
whole account

Check your areas, disk usage and quota: **\$ cindata**

Datamover and Data transfer

Hostname: *data.<clustername>.cineca.it*

Main features

- **No cpu-time limit on processes**
- **Service is containerized and based on restricted GNU shell RUSH**
 - It is not possible to connect directly to the datamover, but the commands must be executed remotely
 - Connection allowed only with valid **SSH certificate** (2FA), no other private/public keys allowed
 - Connection from a CINECA login node does not require SSH certificate (Hostbased authentication enabled)
- **Only few commands are accepted: scp, rsync, sftp, wget, curl**
- **Environment variables are not defined (\$HOME, \$WORK, \$CINECA_SCRATCH)**
 - You have to specify the absolute path location of the files
- **All storage areas of the cluster are visible**

Datamover and Data transfer

Usage examples: listing directories via sftp and copy files

```
$ sftp <username>@data.<cluster_name>.cineca.it:/path/to/be/listed/  
Connected to data.<cluster_name>.cineca.it  
Changing to: /path/to/be/listed/
```

```
sftp> cd /g100_scratch/userexternal/<username>
```

```
sftp> ls -l
```

```
drwxr-xr-x  3 dmolina1 interactive  4096 Mar  3  2023 Speed  
-rw-r--r--  1 dmolina1 interactive 12533510 Oct 12 09:11 test_file
```

```
sftp> get test_file
```

```
Fetching /g100_scratch/userexternal/<username>/test_file to test_file  
test_file                100% 12MB  2.1MB/s  00:05
```

```
sftp> put jobscript.sh
```

```
Uploading jobscript.sh to /g100_scratch/userexternal/<username>/jobscript.sh  
jobscript.sh             100% 2278  19.1KB/s  00:00
```

Datamover and Data transfer

Usage examples: moving data to/from an external machine with rsync or scp

(e.g. your PC, a non-CINECA cluster...)

```
$ rsync -PravzHS /absolute/path/from/file <username>@data.<cluster_name>.cineca.it:/absolute/path/to/
```

```
$ rsync -PravzHS <username>@data.<cluster_name>.cineca.it:/absolute/path/from/file /absolute/path/to/
```

```
$ scp /absolute/path/from/file <username>@data.<cluster_name>.cineca.it:/absolute/path/to/
```

```
$ scp <username>@data.<cluster_name>.cineca.it:/absolute/path/from/file /absolute/path/to/
```

Datamover and Data transfer

Usage examples: moving data between two CINECA clusters with rsync or scp

```
$ ssh -xt <username>@data.<cluster_name_1>.cineca.it rsync -PravzHS /absolute/path/from/file  
<username>@data.<cluster_name_2>.cineca.it:/absolute/path/to/
```

```
$ ssh -xt <username>@data.<cluster_name_1>.cineca.it rsync -PravzHS  
<username>@data.<cluster_name_2>.cineca.it:/absolute/path/from/file /absolute/path/to/
```

```
$ ssh -xt <username>@data.<cluster_name_1>.cineca.it scp /absolute/path/from/file  
<username>@data.<cluster_name_2>.cineca.it:/absolute/path/to/
```

```
$ ssh -xt <username>@data.<cluster_name_1>.cineca.it scp  
<username>@data.<cluster_name_2>.cineca.it:/absolute/path/from/file /absolute/path/to/
```



Thank you