Migrating the INFN-CNAF datacenter to the Bologna Tecnopolo: a status update

Daniele Cesini, Alessandro Cavalli, Andrea Chierici, Vincenzo Ciaschini, Alessandro Costantini, Stefano Dal Pra, Donato De Girolamo, Luca dell'Agnello, Massimo Donatelli, Antonio Falabella, Enrico Fattibene, Francesco Giacomini, Barbara Martelli, Diego Michelotto, Lucia Morganti, Carmelo Pellegrino, Andrea Prosperini, Pier Paolo Ricci, Vladimir Sapunenko, Luigi Scarponi, Antonio Velardo and Stefano Zani

INFN CNAF, v.le B. Pichat 6/2 - 40100 Bologna, IT

Tommaso Boccali

INFN Sezione di Pisa, Largo B. Pontecorvo 3, 56127 Pisa, IT

Lorenzo Chiarelli

INFN CNAF and Consortium GARR, Via dei Tizii 6, 00185 Roma, IT

E-mail: luca.dellagnello@cnaf.infn.it

Abstract. The INFN Tier1 data center is currently located in the premises of the Physics Department of the University of Bologna, where CNAF is also located. During 2023 it will be moved to the "Tecnopolo", the new facility for research, innovation, and technological development in the same city area; the same location is also hosting Leonardo, the pre-exascale supercomputing machine managed by CINECA, co-financed as part of the EuroHPC Joint Undertaking, 4th ranked in the top500 November 2022 list. The construction of the new CNAF data center consists of two phases, corresponding to the computing requirements of LHC: Phase 1 involves an IT power of 3 MW, and Phase 2, starting from 2025, involves an IT power up to 10 MW. The new data center is designed to cope with the computing requirements of the data taking of the HL-LHC experiments, in the time spanning from 2026 to 2040 and will provide, at the same time, computing services for several other INFN experiments and projects, not only belonging to the HEP domain. The co-location with Leonardo opens wider possibilities to integrate HTC and HPC resources and the new CNAF data center will be tightly coupled with it, allowing access from a single entry point to resources located at CNAF and provided by the supercomputer. Data access from both infrastructures will be transparent to users. In this presentation we describe the new data center design, providing a status update on the migration, and we focus on the Leonardo integration showing the results of the preliminary tests to access it from the CNAF access points.

1. Introduction

The National Institute for Nuclear Physics (INFN) is the Italian research agency dedicated to the study of the fundamental constituents of matter and the laws that govern them. To achieve

Year	CPU (kHS06)	Disk (PB)	Tape (PB)
2023	660	70	158
2024	792	83	194
2025	950	100	232
2026	1140	119	279

Table 1. Foreseen increase of resources at CNAF

these goals, it conducts theoretical and experimental research in the fields of sub-nuclear, nuclear and astro-particle physics. All these activities require enormous amounts of computing: since the early 2000s, in anticipation of LHC (Large Hadron Collider), INFN has funded a Tier1 and 9 Tier2s. The main data center, the Tier1, is in Bologna at CNAF, the INFN National Center dedicated to Research and Development on Information and Communication Technologies. The Tier1 currently hosts compute and storage resources for dozens of collaborations beyond those at LHC, while the Tier2s are dedicated to the LHC experiments[1].

The largest part of the installed resources ($\sim 80\%$) at CNAF are dedicated to the experiments at LHC: we expect this to remain true for the coming years as well. Extrapolating the pledges for CNAF according to the WLCG[2] "flat budget" model (20% increase/year) we have the figures reported in Tab. 1 and depicted in Fig. 1.

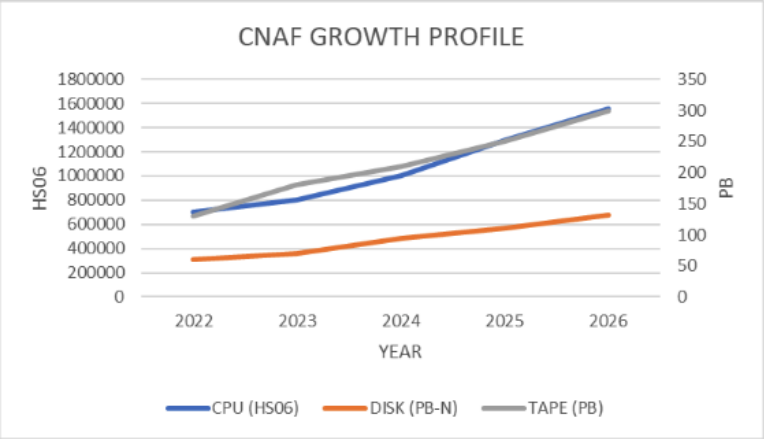

Figure 1. Foreseen increase of resources at CNAF

The CNAF data center could host this expected amount of resources with an upgrade of cooling infrastructure, but it is unlikely, with the current technological trend, the limitations of space and availability of electricity, that it could cope with the huge increase of resources expected for the high-luminosity LHC era (see Fig. 2).

The search for a new location for the data center began in 2017: a combination of events under the aegis of the Emilia-Romagna Region led to the development of a technological district in a brownfield area, the Tecnopolo. The Tecnopolo ($\sim$ 100000 square meters) hosts the new data center of the European Center for Medium-Range Weather Forecasts (ECMWF) [1], while another part has been granted to INFN to host the Tier1 and to the super-computing Italian center CINECA [2] for the pre-exascale machine Leonardo[3].

[1] There are some exceptions such as Bari and Torino Tier2s providing resources to some other collaborations.
[2] Worldwide LHC Computing Grid
[3] Leonardo, funded by the EuroHPC Joint Undertaking, and by the Italian Ministry of Research and University,

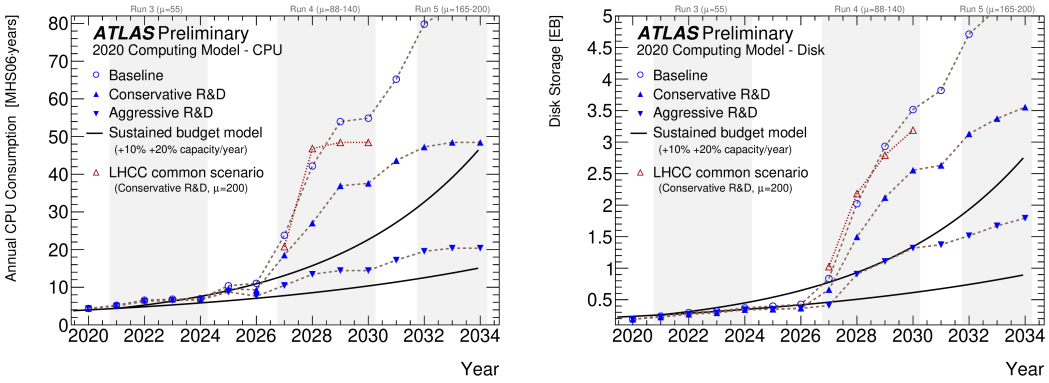

Figure 2. Foreseen increase of CPU (left) and disk (right) resources for Atlas experiment

In the near future, other research entities, such as ICSC[4] [4] [5], and a startups' incubator will be hosted in this district (see Fig. 3).

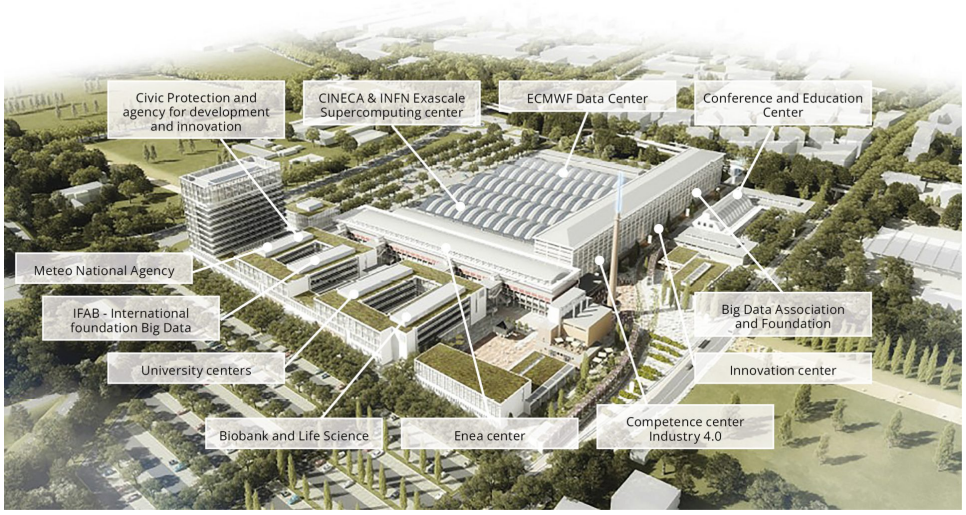

Figure 3. Rendering of the Tecnopolo area.

2. The new data center

The new INFN and CINECA data centers will be housed in two adjacent halls (respectively B5 and C2, see Fig. 4) sharing the technical service infrastructure (i.e., power and cooling). Another building, "Ballette" will possibly host the offices.

While new data center of ECMWF was inaugurated in September 2021 and Leonardo was commissioned in late 2022, the renovation of hall for CNAF data center is expected to be completed by September 2023 (see Fig.5).

will be managed by a consortium composed of CINECA, INFN and SISSA [3]

[4] The High-Performance Computing, Big Data e Quantum Computing Research Centre has been created within the framework of the NextGenerationEU funding by the European Commission.

Figure 4. Areal view of Tecnopolo area.

Figure 5. Front view of B5 during the renovarion works.

While the present data center at CNAF has an usable area of $\sim$ 800 m^2 and a maximum electrical power of 1.4 MW, the new data center at the Tecnopolo will have an usable area for IT larger than 2000 m^2 and it will be possible to ramp up the electrical power from 3 MW in the first phase to 10 MW from 2026 (taking into account also CINECA the total power will ramp up from 16 MW to 25 MW). The target power usage effectiveness (PUE) of the whole data center complex is $\sim$ 1.1.

The hall B5 is divided in 4 main areas, not considering offices and technical rooms (fig. 6):

- Tape library zone (170 m^2), able to host up to 4 tape libraries for a total of 32800 slots (equivalent to $\sim$ 1.5 EB with the newest technology);
- Low density zone (740 m^2) with 128 racks for storage and services (on average, each rack can accommodate resources up to 16 kW): 44 racks will be allocated for the storage (they could accommodate more than 400 PB with the current technology);
- High density zone (200 m^2) able to host up to 3 rows of 14 racks of CPU servers; the use of Direct Liquid Cooling (DLC) will allow to reach a density of 80 $kW/rack$ and hence have more than 6 $MHS06$ of computing power (with the current technology);
- Network zone (117 m^2) for network devices (e.g., access router, core switch) and network

services.

Besides these zones, there is an expansion area (465 m^2) for future needs.

An additional degree of freedom, in the high density zone, is given by the possibility to switch from temperate water (used for DLC) to refrigerated water in the pipes installed under the floating floor. In fig. 7 a view of the high density zone with and without the floating floor is presented (in the latter the pipes for water are visible).

In fact, in the first phase, one of the rows will host racks with Rear-Door Heat Exchangers able to reach a density of at least 40 $kW/rack$ using refrigerated water, while the other two rows will not be used.

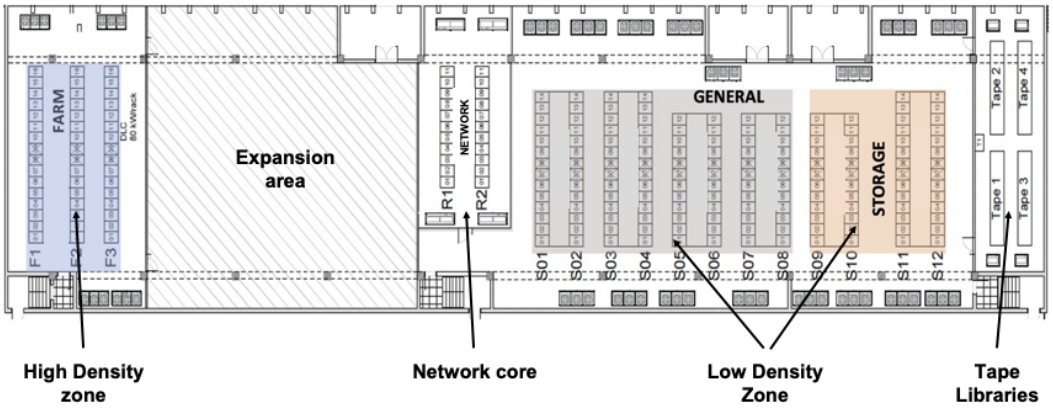

Figure 6. Layout of hall B5

As of September 2023, the renovation works are almost completed and the laying of network cables is about to start.

Figure 7. View of high density zone with (left) and without (right) the floating floor

3. Migration to Tecnopolo

The migration procedure from CNAF to Tecnolopolo is a complex endeavor in itself, but even more complex when considering the zero downtime approach which needs to be obtained for most of the components. Obviously, this cannot be achieved before hall B5 is delivered to CNAF, completely equipped with the secondary distribution of power and cooling, and cabled racks: this should happen by end of October 2023.

The key factors to achieving a zero downtime migration are a high bandwidth link connecting the old and new data centers (1.6 Tbps DCI[5] will be available) and a storage buffer ($\sim$ 60 PB of disk will be installed in advance at the Tecnopolo) in order to copy the data from CNAF before actually moving the storage systems (one by one). All the services will be migrated installing a new instance at the Tecnopolo or, where possible, such as for cloud services, exploiting instance live-migration functionality.

Currently, part of the farm is hosted at CINECA site in Casalecchio: hence during the migration period (we think this will need several months) INFN Tier1 resources will be distributed between Tecncopolo, CINECA and CNAF, with high bandwidth interconnections based on DCI technology.

Moreover, there will be a backdoor to access Leonardo, hosted in the adjacent hall. In fact, Leonardo will provide a non-negligible amount of computing power. In Fig. 8 a schematic view of the network connections is depicted.

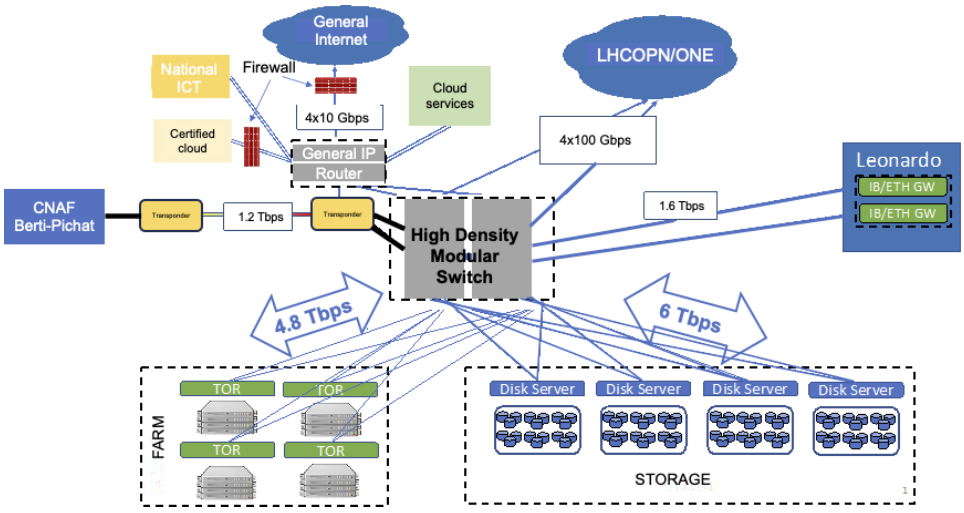

Figure 8. Logical schema of network connections between the new data center at Tecnopolo, CNAF and CINECA.

An additional issue is given by the network topology of Leonardo: this is based on Infiniband technology and the access will be possible only through 2 Infiniband-Ethernet gateways (Mellanox Skyway) with an aggregated bandwidth of 1.6 Tbps. This setup implies several known limitations such as no support for IPv6 and no support for large Ethernet frames[6]. To avoid changing configuration on all disk-servers we configured the Path MTU discovery and tested the following scenario: a rack of computing nodes from the farm configured with MTU=4000 accessing production disk-servers having MTU=9000. The results were promising: no performance degradation or errors were observed (see Fig. 9).

4. Conclusions

The conclusion of the renovation works of the new data center of INFN Tier1 is close (end of October) and, by November 2023, the migration of the resources from the old data center will begin. The challenge will be not only moving the resources and the services without any

[5] Data Center Interconnect (DCI) technology connects two or more data centers together over short, medium, or long distances using high-speed packet-optical connectivity.

[6] The maximum value of MTU (Maximum Transfer Unit) Skyway support is 4092 while on CNAF LAN we have MTU=9000 to maximize the throughput

Figure 9. Results of the simulation of access to storage through the Skyway systems.

downtime, but, at the same time, also to be able to manage a data center distributed over three sites and efficiently use the pre-exascale machine Leonardo.

5. Acknowledgements

This work is partially supported by ICSC - the High-Performance Computing, Big Data e Quantum Computing Research Centre, funded by European Union – NextGenerationEU.

References

[1] [Online]. Available: `https://ecmwf.int`
[2] [Online]. Available: `https://www.cineca.it/en`
[3] [Online]. Available: `https://www.sissa.it/`
[4] [Online]. Available: `https://commission.europa.eu/strategy-and-policy/recovery-plan-europe_en`
[5] [Online]. Available: `https://www.supercomputing-icsc.it/en/icsc-home/`