Challenges and opportunities in migrating the CNAF data center to the Bologna Tecnopolo

Daniele Cesini, Alessandro Cavalli, Andrea Chierici, Vincenzo Ciaschini, Alessandro Costantini, Stefano Dal Pra, Donato De Girolamo, Luca dell'Agnello, Massimo Doantelli, 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: daniele.cesini@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. Soon, it will be moved to the "Tecnopolo", the new facility for research, innovation, and technological development in the same city area; it will follow the installation of Leonardo, the pre-exascale supercomputing machine managed by CINECA, co-financed as part of the EuroHPC JU (Joint Undertaking). The construction of the new CNAF data center will consist of two phases, corresponding to the computing requirements of LHC: Phase 1, starting from 2023, will involve an IT power of 3 MW, and Phase 2, starting from 2026, involving an IT power up to 10 MW. The primary goal of the new data center is to cope with the computing requirements of the data taking of the HL-LHC experiments, in the timeframe spanning from 2026 to 2040, providing, at the same time, computing services for several other INFN experiments, projects, and activities of interest, either they are currently in operation, under construction, in advanced design, or even not yet defined. The co-location with Leonardo will open new scenarios, with a close integration between the two systems able to share dynamically resources. In this contribution we describe the new center design, with a particular focus on the status of the migration, its schedule, and the technical challenges we will face moving the data center without service interruption. On top of this, we will analyse the opportunities that the new infrastructure will open in the context of the NRRP (National Resilience and Recovery Plan) funding and strategic plans, within and beyond the High Energy Physics domain.

1. Introduction

CNAF, the national center of INFN (Italian Institute for Nuclear Physics) dedicated to Research and Development on Information and Communication Technologies in Bologna, currently hosts the largest computing center deployed by INFN, and it is part of a federation of other 9 (smaller) INFN data centers initially created to satisfy the needs of the LHC experiments at CERN; in the WLCG hierarchy, CNAF is certified as the only Tier1 center in Italy.

CNAF, with the current IT technology, would be able to host IT resources to cover the requirements up to the end of LHC Run 3 (fig. 1) which, following the WLCG "flat budget" model (20% increase/year) applied to the 2023 CNAF pledges (also including the needs of the other experiments) will be:

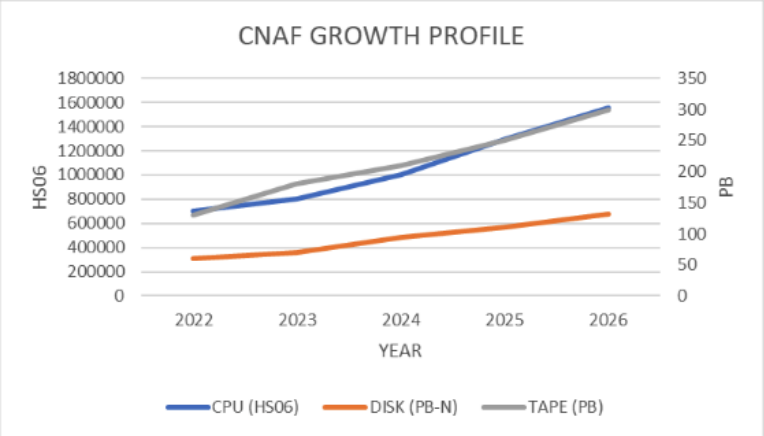

Figure 1. Foreseen increase of resources at CNAF

- HTC computing power: 1,2 MHS06
- Storage on disk: 100 PB
- Storage on Tape: 230 PB

This amount of resources still would fit in the infrastructure of the data center, but, as early as 2017, it was clear that a new data center would be needed to accommodate the expected huge increase in resources needed for the High-Luminosity LHC era (see fig. 2).

2. The Data Valley district

In the year 2017, the Emilia-Romagna Region decided to promote a new district, devoted to research, innovation and technological development, redeveloping a disused area ($\sim$ 100000 square meters), formerly a tobacco factory.

The main purpose of this project is to create a large scientific center of excellence in the field of high-performance computing which also acts as a hub for technological development. In the framework of this operation, part of the area has been granted to hosting 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 its new data center (the Tier1) and to CINECA [2] for Leonardo, one of the three pre-exascale machines funded by the EuroHPC Joint Undertaking, and by the Italian Ministry of Research and University. Leonardo, with a peak computing power of over 300 Petaflop/s, will be managed by a consortium composed of CINECA, INFN and SISSA [3]. In the near future, other research entities and a startups' incubator will be hosted in the Data Valley district (see Fig. 3).

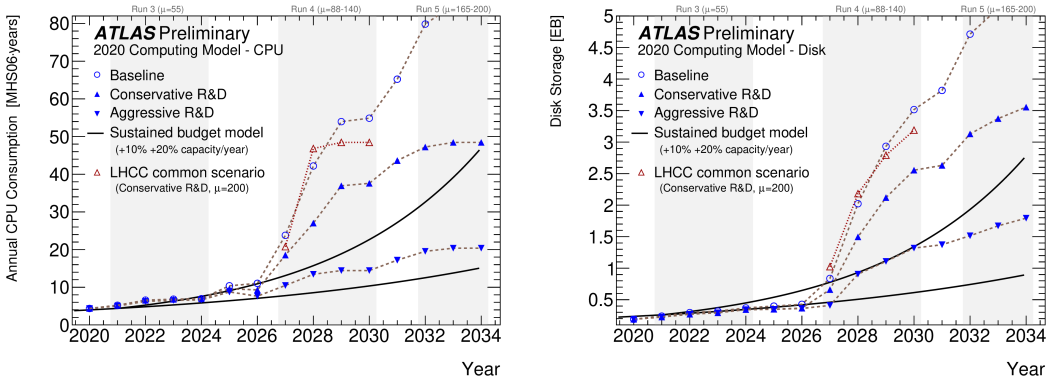

Figure 2. Foreseen increase of CPU (left) and disk (right) resources for Atlas experiment

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 July 2023.

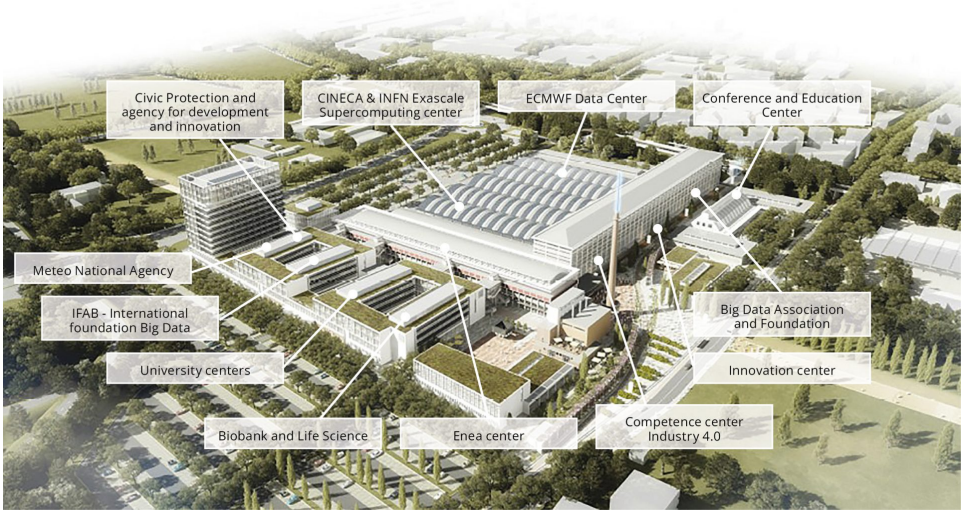

Figure 3. Rendering of the Tecnopolo area.

3. CNAF and the Tecnopolo

CNAF in its new location at the Data Valley district is expected to maintain or even raise its role within the INFN distributed computing landscape. In fact, the transfer of the data center is also a great opportunity to evolve by embracing state-of-the-art technologies in many areas, and to provide a sustainable path to support the evolution of computing services and infrastructures within INFN, starting from the challenge to manage a technical service infrastructure shared with CINECA.

The primary goal of the new data center is to cope with the computing requirements of the data taking of the HL-LHC experiments, within the horizon 2026 to 2040, providing at the same time computing services for many other INFN experiments, projects, and activities of INFN

interest, be they currently in operation, under construction, in advanced design, or even not yet defined.

3.1. A new data center for CNAF

The new INFN and CINECA data centers will be housed in two adjacent halls (respectively B5 and C2) and they will share the technical service infrastructure (i.e., power and cooling) .

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 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 13 MW to 25 MW).

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

- Tape library zone (170 m^2), able to host up to 4 tape libraries for a total of 32800 slots (equivalent to $\sim 1\ EB$ with forthcoming 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 the Direct Liquid Cooling will allow to reach a density of 80 $kW/racks$ 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.

As of February 2023, the renovation works are still on going: in fig. 5 a view of the high density zone with pipes for Direct Liquid Cooling and one of the low density zone is presented (the floating floor is still to be installed).

4. The road to the Tecnopolo: the services

The definition of the baseline plan for CNAF migration has required intense R&D activities, in order to evaluate the diverse solutions available from technology tracking or from internal development. In fact, the final goal, in addition to moving the hardware from the old data center to the new one, is also a complete renewal of the services. Both these aspects have been addressed in a Technical Design Report and the related migration plan.

The standard services provided by the Tier1 are the batch system (currently provided by an HTCondor instance) and a storage archive (with both classes of service online and near-line). Recently, there have been requests for a cloud service from some communities: to address this requirement, INFN has launched the INFN cloud initiative to build a corporate cloud. So one of the major challenges we are facing is defining a strategy for integrating data centers into this new cloud infrastructure. The CNAF Cloud infrastructures - based on the OpenStack framework and providing access to both IaaS and PaaS levels - is already federated with the national INFN-Cloud. CNAF started an R&D activity path towards the unification of the HTC and Cloud infrastructures to provide easy migration of resources from one to the other according to the user or administrators needs and avoiding static partitioning. This activity implies a deep reorganization of the datacenter LAN and of the OpenStack network nodes that, at the moment of writing, is still in progress and will be completed before the migration to the Tecnopolo.

Also for the storage component, the natural evolution in a cloud landscape is the integration of all the sites in a Data Lake. This is also the strategy endorsed by WLCG to cope with the

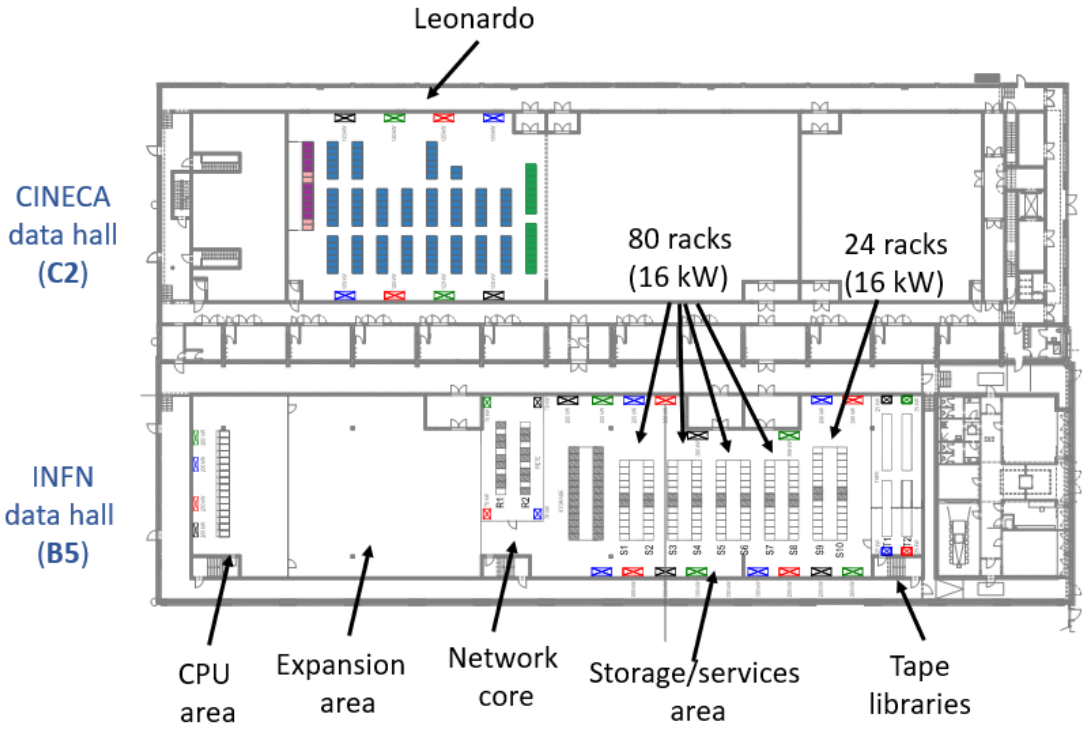

Figure 4. Layout of halls B5 and C2

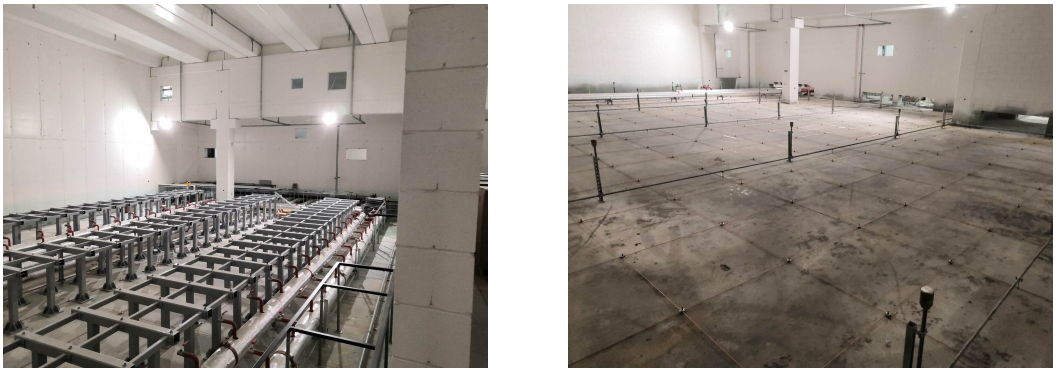

Figure 5. View of high density (left) and low density (right) zones

HL-LHC challenges and it should allow for a cost reduction in the provisioning and maintenance of storage systems. The Tier1 will be the main site composing the Italian Data Lake node, with a few other big storage sites. A key element will be the interconnection at very high bandwidth (multiples of 100 Gbit/s). The networking technologies most suited for this implementation, together with the software solutions used to create, manage and access Data Lakes for huge collaborations (WLCG, ESCAPE) are under investigation, i.e., for the transparent creation of redundant federated storage resources, automatic replication in different, geographically distributed, availability zones and so on.

On the other hand, at local level at CNAF, the bulk of data centre storage is based on robust and stable (yet commercial) solution (Spectrum Scale a.k.a. GPFS, together with Spectrum Protect a.k.a. TSM, both by IBM, for the archival part). Both to overcome the cost of GPFS

licenses and to better address some cloud use cases, an open source alternative, CEPH, has been tested and validated: it is currently in production as storage system for the Alice experiment and for some small collaborations using the cloud. Moreover, the national infrastructure for the users' home directories, currently under development, will be based on CEPH.

Another key component of the infrastructure is the AAI system: After an R&D phase involving existing authorization systems on the market, FreeIPA/Keycloak with INDIGO-IAM were chosen as the new authorization system, driven by the specs of the tools and by the results of evaluations done within the HEPiX community. Several other data centres like ours decided to adopt these tools (among them, CERN). The combination of FreeIPA, Keycloak and INDIGO-IAM will provide a flexible and modern AAI solution giving CNAF the opportunity to implement key security procedures tailored on our customers' requirements, like, e.g., multi-factor authentication, and simplifying day-by-day administration activities.

5. Migration plan 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.

The key factors to achieving a zero downtime migration are a high bandwidth link connecting the old and new data centers (1.6 Tbps link will be available) and a storage buffer (32 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 and the feature of CEPH cluster to rebalance data during osd node maintenance.

6. Conclusions

In the present paper, the activities aimed at moving the INFN Tier1 data center to the "Tecnopolo", the new facility for research, innovation, and technological development in the same city area, is described and reported, Co-location with Leonardo, the new INFN Tier1 data center will be ready to face new scenarios, thanks to a close integration between the two systems and to the ability to share dynamically resources, and also drive the new technological challenges we will have to deal with in the coming years. In this contribution the new center design, the status of the migration, its schedule, and the technological plans have been reported, together with the design and the definition of the technical design report and the migration plan, that will be carried out without any service interruption.

References

[1] [Online]. Available: `https://ecmwf.int`
[2] [Online]. Available: `https://www.cineca.it/en`
[3] [Online]. Available: `https://www.sissa.it/`