EPJ Web of Conferences **214**, 03001 (2019) https://doi.org/10.1051/epjconf/201921403001
CHEP 2018

The INFN scientific computing infrastructure: present status and future evolution

T. Boccali[1], *G.* Carlino[2], and *L.* dell'Agnello[3,*] on behalf of the C3S Committee

[1]INFN Sezione di Pisa, Largo B. Pontecorvo 3, 56127 Pisa, Italy
[2]INFN Sezione di Napoli, via Cintia, 80126 Napoli, Italy
[3]INFN CNAF, v.le B. Pichat 6/2, 40100 Bologna, Italy

Abstract. The INFN scientific computing infrastructure is composed of more than 30 sites, ranging from CNAF (Tier-1 for LHC and main data center for nearly 30 other experiments) and nine LHC Tier-2s, to ~ 20 smaller sites, including LHC Tier-3s and not-LHC experiment farms. A comprehensive review of the installed resources, together with plans for the near future, has been collected during the second half of 2017, and provides a general view of the infrastructure, its costs and its potential for expansions; it also shows the general trends in software and hardware solutions utilized in a complex reality as INFN. As of the end of 2017, the total installed CPU power exceeded 800 kHS06 (~ 80,000 cores) while the total storage net capacity was over 57 PB on disk and 97 PB on tape: the vast majority of resources (95% of cores and 95% of storage) are concentrated in the 16 largest centers. Future evolutions are explored and are towards the consolidation into big centers; this has required a rethinking of the access policies and protocols in order to enable diverse scientific communities, beyond LHC, to fruitfully exploit the INFN resources. On top of that, such an infrastructure will be used beyond INFN experiments, and will be part of the Italian infrastructure, comprising other research institutes, universities and HPC centers.

1 Introduction

The National Institute for Nuclear Physics (INFN) is the research agency, funded by the Italian government, dedicated to the study of the fundamental constituents of matter and the laws that govern them. It conducts theoretical and experimental researches in the fields of sub-nuclear, nuclear and astroparticle physics. Since its foundation, in 1951, the activities it fosters require larger and larger computing and storage resources, due to the increasing complexity of the experiments. INFN has also been the seminal institution for the national research network in Italy, now handled by GARR[1].

INFN has 20 divisions, four national laboratories and two national centers: most of them host one computing center at least, varying in size from the WLCG Tier-1[2] to small local facilities.

*e-mail: luca.dellagnello@cnaf.infn.it

[1]https://www.garr.it/en/

[2]For the role of Tier-0, Tier-1 and Tier-2 centers in the computing of the experiments at LHC see for example http://wlcg-public.web.cern.ch/tier-centres

Table 1: INFN computing infrastructure (end of 2017)

Resource Type	Unit of Measure	Value (2017)
CPU	Cores	70,000
Disk	TB-N	57,000
Tape	TB	97,500
Electric Power	MW	3.4
Staff	FTE	92*

A new committee, C3S[3], has been formed two years ago to coordinate the scientific computing at INFN and lead its evolution.

As a first step, in order to get a complete picture to plan future evolution and find possible optimizations in the short term, the C3S organized a survey of the INFN computing centers. We will present, in the next sections, the results of the survey and identify some possible evolutionary scenarios.

2 Cost, expandability and optimization

The focus of the survey was on the resource deployment (figures relative to the end of 2017), the infrastructure capabilities and the effort needed for their support. For this reason, only a questionnaire was requested for each computing infrastructure, even if hosting several logical data centers (such as CNAF with the INFN Tier-1 and a Tier-2 for the LHCb experiment[4]); on the other hand, for other sites, e.g. Napoli, more questionnaires were needed due to multiple infrastructures present. The most relevant aggregate values for the infrastructure are reported in Tab. 1.

We collected 29 distinct answers, for data centers ranging from few tens to more than 15,000 hosted CPU cores. In order to ease the analysis of the collected data, we divided the sites into three categories:

- Large size data center - having a computing farm with more than 1,000 cores and installed disk of at least 750 TB-N[5]. These values correspond to $\sim$ 50% of the amount of available resources at the average INFN Tier-2 in 2017;
- Small size data center - having a computing farm with less than 200 cores or less than 100 TB-N of disk[6];
- Medium size data center - all the remaining centers (8 in total).

As shown in Fig. 1, with the chosen thresholds, the three groups are clearly identified and separated.

[3] *Comitato di Coordinamento del Calcolo Scientifico* (Scientific Computing Coordination Committee)
[*] $\sim$ 32% is non INFN personnel, usually from a cohosting university department.
[4] The *Large Hadron Collider beauty* experiment at CERN (http://lhcb-public.web.cern.ch/lhcb-public/)
[5] TB-N or net TB indicates the usable capacity, expressed in TeraBytes, after the RAID and file-system overhead.
[6] These resources could be hosted in a single rack unit and use less than 2 kW of power

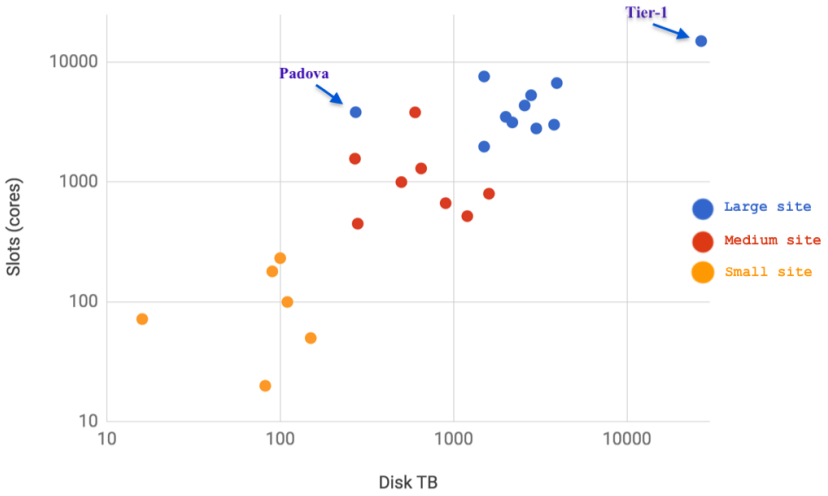

Figure 1. Cores vs Disk distribution at INFN.

Table 2: INFN sites grouped by size

Large size sites	**Medium size sites**	**Small size sites**
CNAF	Napoli	Milano Bicocca
Torino	Roma3	LNS
Pisa	Genova	Pavia
Milano	Consenza	Roma2 (2 data centers)
LNL	Trieste	Perugia
Padova	LNF (KLOE)	Ferrara
Roma1 (Tier-2)	Parma	Roma1 (non Tier-2)
LNF (Tier-2)	LNGS	Firenze
Bari		
Napoli (Tier-2)		
Catania		

Eventually, the large sites correspond to the official WLCG Tier-1 and Tier-2 sites[7] as depicted in Tab. 2. The division of sites into these three categories helps us in understanding the different cost patterns: as probably expected, there is a higher need of support for small installations (when scaled by the resource deployment, see Fig. 2 and Fig. 3). In fact, the cost of support per unit resource is evidently smaller for large sites, which suggests (or better, confirms) that economies of scale are important.

Another important economic aspect is the cost of electric power: for the Tier-2s hosted by a University department, it is so far included in the rental agreement (€2.4M is the annual "in-kind" contribution evaluated with current power costs). Although this is a clear advantage at the moment for INFN, it is uncertain whether this could be true for the future, because of the increase in the amount of the resources foreseen in the coming years.

[7]The INFN Padova data center has been included in the large sites, since it is part of the INFN Legnaro Tier-2, via a dedicated connection.

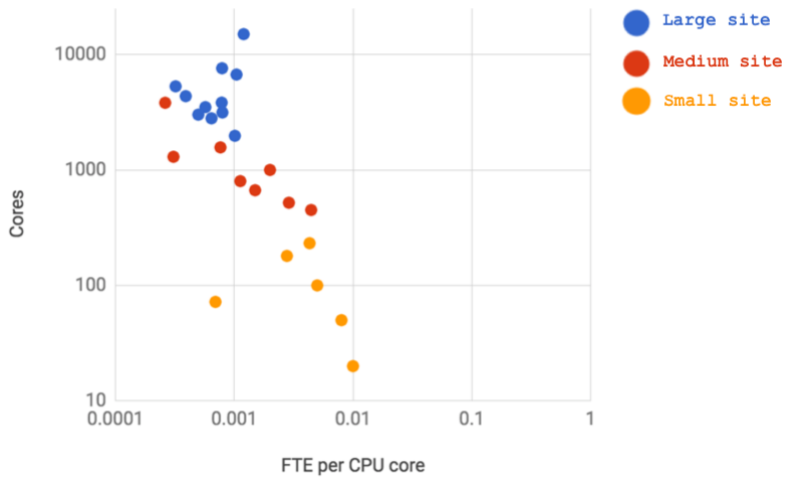

Figure 2. FTE needed per core.

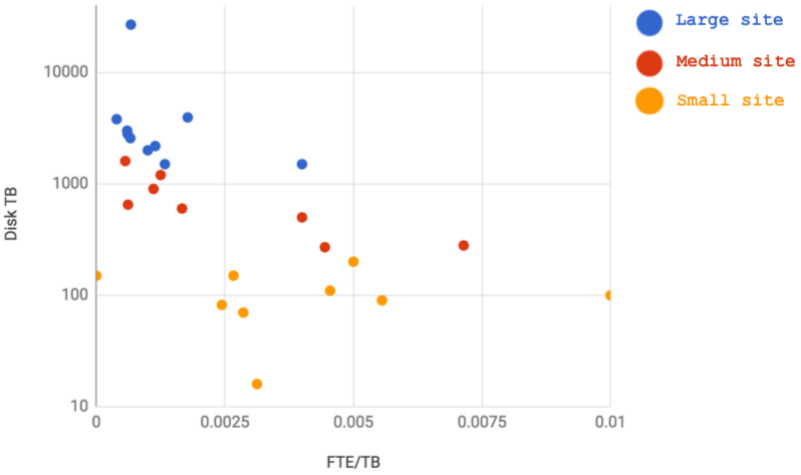

Figure 3. FTE needed per TB-N.

Concerning the capability for expansion, it was found that a +60% increase in space, power and cooling can be accommodated with current INFN distributed data centers, without the need of major infrastructure reworkings. Anything larger requires the deployment of different centers. Still, assuming the validity of Moore's law at a level of +20%/year, this means the current infrastructure could sustain a deployment of resources seven times larger by 2026.

3 Future directions

The need for scientific computing is expected to increase widely in the next decade. High Luminosity LHC (HL-LHC) projects a 20x increase in needs by 2027, which cannot be ac-

commodated in the current infrastructure. By the same date, other experiments like CTA [1] and SKA [2] are expected to require the nearly same amount of resources as a typical LHC experiment.

To cope with these challenges, INFN has started several R&D activities towards HL-LHC, mainly through the participation to European projects:

- Evolution of Computing towards Clouds with the European projects INDIGO-DataCloud [3] for the development of the services, HNSciCloud [4] to test the use of commercial clouds for the research, EOSCpilot [5] and EOSC-hub [6] to build the European Open Science Cloud;
- Study of solutions for Exabyte level storage, including caches and optimized access through the European project eXtreme Data Cloud (XDC) [7]);
- Test of use of GPUs via Cloud Interfaces with the European project DEEP-Hybrid DataCloud (DEEP) [8];
- Study of 'Data Lakes' via the European project ESCAPE (European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures).
- Test of a common Computing and Data Infrastructure lake among data centers in the south of Italy in the framework of the PON[8] IBiSCo, still under evaluation. The project goal is to carry out the strengthening of the current southern ReCaS [9] computing infrastructure, already funded with INFN ordinary funding, past PON programs and the Distributed High Throughput Computing and Storage in Italy (DHTCS-IT) project also funded by the Italian Ministry of Education, and to constitute the first real and concrete step towards the Italian Computing and Data Infrastructure, a multi-disciplinary and multi-functional platform, which is able to adapt to the needs of all the scientific communities.

Furthermore, INFN is exploring unconventional ways to increase the computing served by its Tier-1 center at CNAF, via cloud and remote resources: tests were performed using commercial cloud providers (Aruba, Azure, T-Systems), academic sites (Bari-ReCaS), and more recently by offloading a large fraction of CNAF computing resources to the PRACE Tier-0 at CINECA [10].

These R&D programs enable INFN to disentangle a variety of issues like:

- The need for caching systems to overcome latency and bandwidth limits (solutions like GPFS/AFM [11] and Xcache [12] have been tested on production systems).
- The need to be able to cope with different cloud technologies (VMWare[9], Azure[10], OpenStack[11], OpenNebula[12]); in this respect, INFN has been the promoter of the Dynamic On Demand Analysis Service (DODAS) [13], which is currently being tested in several CMS[13] centers (e.g. opportunistic Tier-3 sites on commercial cloud providers such as Azure and regular Tier-2 sites such as Imperial College, IFCA[14] and Sofia);
- The need of high bandwidth connectivity among remote sites: the experimentation had started with the distributed Tier-2 between INFN-Legnaro and INFN-Padova, and a 1.2 Tbit/s link is now operational between CNAF and CINECA via the emerging technology of Data Center Interconnect (DCI) [10].

[8] "National Operative Program - Research and Innovation 2014-2020" funded by Italian Ministry of Education
[9] http://www.vmware.com/
[10] https://azure.microsoft.com/
[11] https://www.openstack.org/
[12] https://opennebula.org/
[13] The *Compact Muon Solenoid* experiment at CERN (https://cms.cern)
[14] Institute of Physics of Cantabria (https://ifca.unican.es/es-es/project?exp=25)

- The need to incorporate external computing entities like HPC sites into the INFN infrastructure, either via opportunistic allocations, or grants. The handshaking with HPC sites is generally difficult, due to site policies (access, networking, local disk space, ...), and the experimentation allows to gain experience on how to develop and deploy custom solutions.

4 Towards the Data Lake

In the framework of XDC, ESCAPE and collaborations with other European centers, INFN is performing tests on possible Data Lake solutions, which are currently the most popular options for the storage handling in the 2020s. INFN has also launched a specific R&D experiment, the Italian Distributed Data Lake for Science (IDDLS) initiative, in order to test "Data Lake like" solutions by linking three WLCG sites in Italy, with a mesh of DCI solutions.

A great opportunity is given from the new facility of European Centre for Medium-Range Weather Forecasts (ECMWF) [14] which is going to be launched in Bologna from 2020. This facility will accommodate not only a data center for ECMWF but also another one for CINECA and CNAF. On one hand, the co-location of CINECA and CNAF will allow for additional and closer interactions between the INFN HTC and the PRACE HPC infrastructures; on the other hand, the new data center, with a dedicated power up to 10 MW for INFN computing and data resources, will constitute the main component of the future INFN Data Lake.

As mentioned in Sec. 3, also in the program of the project IBiSCo it is foreseen the setup of a Data Lake among the main sites in the south of Italy.

The structure of a Data Lake in Italy is still an open question with at least three variants (see Fig. 4):

- The main Data Center in Bologna, other sites serving as Compute Nodes, possibly with caches;
- A single logical Data Center, physically distributed with co-location of CPUs and storage;
- Compute nodes logically distributed per region.

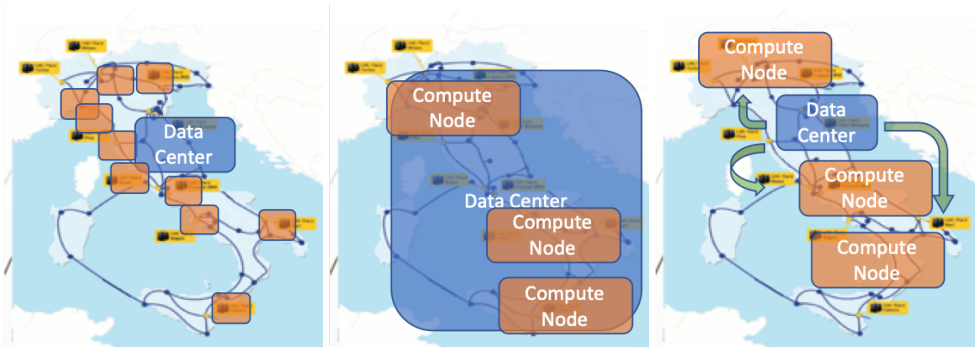

Figure 4. Possible Data Lake models for INFN.

While the centralized approach represented in the first and third scenarios (ie a single Data Center hosting the data, see Fig. 4) could help in the optimization of resource management and would fit into a purely WLCG ecosystem (redundancy would be provided by other centers at European level), on the other hand, it would not be optimal for other experiments that could

instead require some kind of data redundancy within the INFN infrastructure. Furthermore, the consolidation of the present Tier-2 infrastructure in fewer larger sites could reduce the total cost of its maintenance in the medium to long term. Thus, a probable scenario for the Data Lake could see a distributed site for data, with the main data center in Bologna and a smaller one in the South for redundancy of the data not replicated elsewhere, and a few large sites with computing nodes (see the middle picture of Fig. 4) .

5 Conclusions

The facility under construction in Bologna for the new ECMWF data center will also be able to host the CNAF and CINECA data centers. The new CNAF data center will have a dedicated power of up to 10 MW for IT and will constitute the main component of the future INFN Data Lake.

The final deployment of INFN computing for the next decade is still not decided, but most probably it will involve:

- the consolidation of storage in fewer sites;
- the optimization of the Tier-2 infrastructure into bigger sites;
- the exploitation of the coming Tier-1 data center and possibly of another one that would be constructed in the south of Italy (the follow-up of ReCaS);
- the realization of a closer connection of HTC and HPC resources, enhanced by the colocation of CNAF and CINECA;
- the consolidation of small sites into the INFN wide Cloud infrastructure.

References

[1] S. Vercellone *for the CTA Consortium*, *"The next generation Cherenkov Telescope Array observatory: CTA"*, Nucl. Instrum. Meth. A **766** (2014) doi: 10.1016/j.nima.2014.04.015

[2] https://www.skatelescope.org/

[3] D. Salomoni *et al*, *"INDIGO-DataCloud: a Platform to Facilitate Seamless Access to E-Infrastructures"*, J. of Grid Computing **16** (2018) doi: 10.1007/s10723-018-9453-3

[4] M. Gasthuber *et al*, *"HNSciCloud - Overview and technical Challenges"* J. Phys.: Conf. Ser. **898**, 052040 (2017) doi: 10.1088/1742-6596/898/5/052040

[5] https://eoscpilot.eu/media/publications

[6] https://www.eosc-hub.eu/publications

[7] D. Cesini *et al*, *"The eXtreme-DataCloud project: data management services for the next generation distributed e-infrastructures"*, Proceedings of Conference Grid, Cloud & High Performance Computing in Science (ROLCG), 1-4 (2018) doi: 10.1109/ROLCG.2018.8572025

[8] https://deep-hybrid-datacloud.eu/

[9] G. Andronico *et al*, *"High Performance Scientific Computing Using Distributed Infrastructures"* (World Scientific, 2017), doi: 10.1142/9972

[10] L. dell'Agnello *et al*, *"INFN-Tier1: a distributed site"*, Proceedings of CHEP 2018, to be published

[11] T. Boccali *et al*, *"Extending the farm on external sites: the INFN Tier-1 experience"*, J. Phys.: Conf. Ser. **898**, 082018 (2017) doi: 10.1088/1742-6596/898/8/082018

[12] L. A. T. Bauerdick *et al*, *"XRootd, disk-based, caching proxy for optimization of data access, data placement and data replication"*, J. Phys.: Conf. Ser. **513**, 042044 (2014) doi: 10.1088/1742-6596/513/4/042044

[13] D. Spiga *et al*, *"DODAS: How to effectively exploit heterogeneous clouds for scientific computations"*, PoS ISGC 2018 & FCDD, 024 (2018) doi: 10.22323/1.327.0024

[14] https://www.ecmwf.int/en/about/media-centre/news/2017/ecmwfs-new-data-centre-be-located-bologna-italy-2019