EPJ Web of Conferences **214**, 09008 (2019) https://doi.org/10.1051/epjconf/201921409008
CHEP 2018

Disaster recovery of the INFN Tier-1 data center: lesson learned

Luca dell'Agnello on behalf of all CNAF colleagues[1,*]

[1]INFN-CNAF, v.le B. Pichat 6/2 - 40100 Bologna, Italy

Abstract. The year 2017 was most likely a turning point for the INFN Tier-1. In fact, on November 9th 2017 early at morning, a large pipe of the city aqueduct, located under the road next to CNAF, broke. As a consequence, a river of water and mud flowed towards the Tier-1 data center. The level of the water did not exceed the threshold of safety of the waterproof doors but, due to the porosity of the external walls and the floor, it could find a way into the data center. The flooding almost compromised all the activities and represented a serious threat to future of the Tier-1 itself. The most affected part of the data center was the electrical room, with all switchboards for both power lines and for the continuity systems, but the damages were diffused also to all the IT systems, including all the storage devices and the tape library. After a careful assessment of the damages, an intense recovery activity was launched, aimed not only to restore the services but also to secure data stored on disks and tapes. After nearly two months, in January, we were able to start to reopen gradually all the services, including part of the farm and the storage systems. The long tail of recovery (tapes recovery, second power line) has lasted until the end of May. As a short term consequence we have started a deep consolidation of the data center infrastructure to be able to cope also with this type of incidents; for the medium and long term we are working to move to a new, larger, location, able also to accommodate the foreseen increase of resources for HL-LHC.

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. The INFN is composed by more than 20 divisions dislocated at the main Italian University Physics Departments, 4 Laboratories and 3 National Centers dedicated to specific tasks (Fig. 1).

CNAF is the National Center of the INFN "for the Research and Development in Information and Communication Technologies": it participated as a primary contributor in the development of Grid middleware and then in the operation of the Italian Grid infrastructure.

2 The CNAF data center

The CNAF data center has a total area of about 1,000 square meters. It is located two levels under the street. It is divided in 4 main halls (Fig. 2): 2 halls for the IT equipment, 1 small hall for the GARR[1] Point of Presence and 1 electrical room.

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

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (http://creativecommons.org/licenses/by/4.0/).

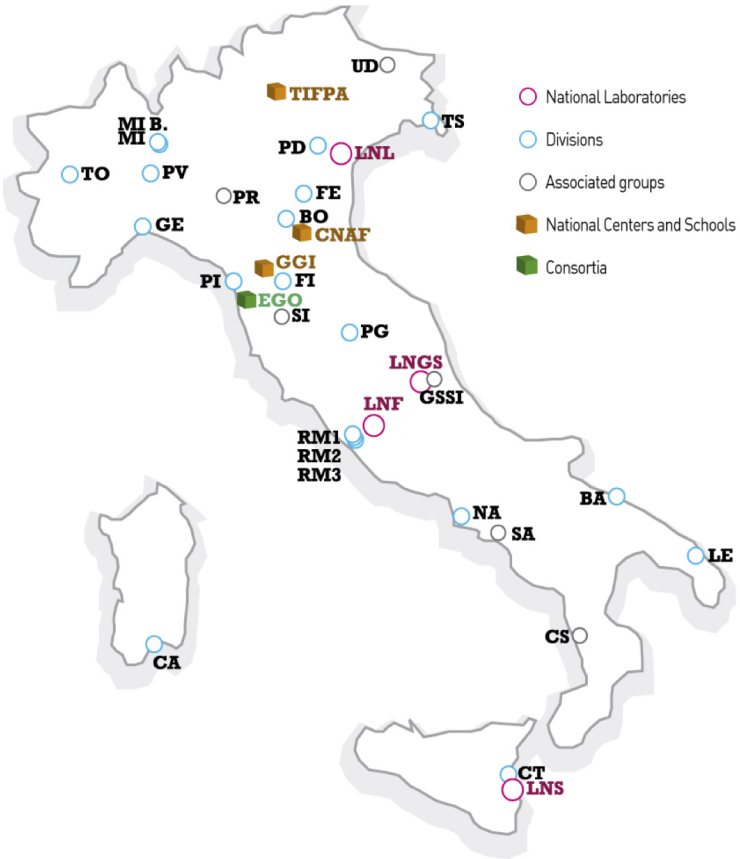

Figure 1. Dislocation of the INFN divisions

The first incarnation of the CNAF data center dates back to 2003 as computing facility for the BaBar[2], CDF[3] and Virgo[4] experiments and also as a prototype for the INFN Tier-1 for the future high-energy physics experiments at the Large Hadron Collider (LHC) in Geneva: ALICE, ATLAS, CMS, and LHCb.

In 2008 the data center underwent a complete infrastructural refurbishing in order to allow to host up to 1.4 MW of IT resources and have a full redundancy for power (2 independent lines) and cooling. Nowadays, besides Virgo and the four experiments at LHC, the INFN Tier-1 provides the resources, support and services needed for all the activities of data storage and distribution, data processing, Monte Carlo production and data analysis to about another 30 scientific collaborations, including Belle II and several astro-particle experiments.

Before the flooding, the data center hosted a computing farm composed by ~ 1,000 servers (or WNs) for a total of ~ 20,000 computing slots and a power of ~ 220 kHS06 (with additional ~ 20 kHS06 dislocated in the ReCaS[5] facility in Bari). Also, a small (~ 33 TFlops) HPC cluster was available for specific applications.

The data produced by the experiments were stored on ~ 23.4 PB of disk and in a tape library loaded with ~ 6,000 tapes (for a total of ~ 42 PB of data).

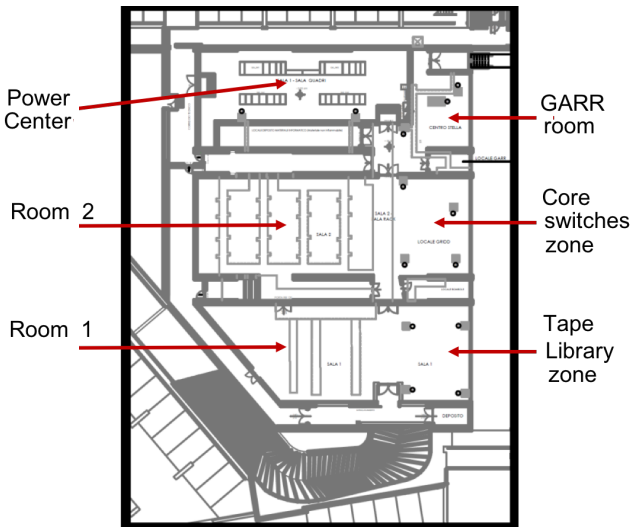

Figure 2. Schematic view of the data center

At the end of 2017, a total of 21 people were working at the Tier-1 (including facilities support staff).

3 The flooding

On 9 November 2017, early at the morning, an aqueduct pipe located in the street nearby CNAF, broke. As a result, a river of water and mud flowed towards the Tier-1 data center (Fig. 3). The water level did not exceed the threshold of safety of the waterproof doors (Fig. 4) but, due to the porosity of the external walls and the floor, it could find a way into the data center.

Both electric lines failed at about 7:10 AM CET. Access to the data center was possible only in the afternoon, after all the water had been pumped out. This operation was not easy: the first attempt failed due to the road collapsing under the weight of the fire truck (Fig. 5 and 6).

In the end, about $500m^3$ of water had entered the data center. While in the IT halls the level of the water reached the lower 4 units of each rack ($\sim$ 30 cm), in the power center, located at a lower level, it reached half of the height of the power equipment ($\sim$ 1 m), thus causing a black-out. In Fig. 7 the water level during the draining operations can be seen.

4 Consequences

Since the four lower units of all racks in the IT halls and the two lower rows of tapes in the library were submerged, nearly all CNAF services suffered serious damages.

In particular, all storage systems were involved for a total $\sim$ 4 net PB compromised. However, it's worth to notice that the impact on each type of system was different depending on its architecture (Distributed Raid over a horizontal crate versus RAID 6 protection with vertical stripes). See Table 1 for a summary of the affected storage systems.

Figure 3. 1) The data center is located under the red area. 2) Relative position of the electric room. 3) From this point, the water poured into the campus from below the street.

Table 1. Effect of the flooding on the experiments data.

#	System	Phase-out year	Affected disks (enclosures)	Protection method	Status	Involved experiments
4	DDN S2A9900	2017	240 (4)	RAID 6 (8+2)	Degraded	ALICE, ATLAS, LHCb
1	DDN SFA10K	2017	120 (2)	RAID 6 (8+2)	Degraded	LHCb
1	DDN SFA12K	2018	168 (2)	RAID 6 (8+2)	Degraded	CMS
2	Dell MD3860f	2020	48 (2)	Distributed RAID	At risk	Darkside, Virgo
1	DDN SFA10K	2021	84 (1)	RAID 6 (8+2)	Degraded	ALICE, ATLAS, AMS
1	DDN SFA10K	2021	168 (2)	RAID 6 (8+2)	No parity	
1	Huawei 6800v3	2022	150 (2)	Distributed RAID	At risk	All non-LHC experiments excepting Darkside, Virgo

The water damaged also the tape library (1 drive and several other components) and contaminated almost 160 tapes (some of these empty, Table 2 for details) in it. Also, the majority of tapes with data from CDF Run 1 was compromised. Also some ~ 14% of the computing power of the farm was lost. On the other hand, the 3 Core Switch/Routers and the General IP Router were safe for few centimeters.

Figure 4. The entrance to the data center on November 9 early morning

Figure 5. The fire truck being extracted from the pothole

Figure 6. The entrance of CNAF where the road collapsed

.

5 Recovery

Besides the IT resources, the most compromised part was the power center (it is located in the lowest part of the data center): both the power lines were lost (including the control for UPS's/diesel engines).

So, the first mandatory operation has been to activate a temporary power line in order to be able to dry the data center. After that, it became possible to start the cleaning of dust and mud: this operation was completed, by a specialized company, during the first week of December. Just before Christmas, the recovery of one of the power lines was completed: to supply the lack of the continuity system, restored only later in mid-February, a temporary UPS to provide continuity for the network and storage systems was leased.

Figure 7. The power room

Table 2. Wet tapes

Experiment	# of wet tapes
ALICE	22
ATLAS	26
CMS	42
LHCb	30
Argo	1
CDF	16
CDF (Run 1)	651
Icarus	2
Kloe	2
Pamela	2
Virgo	4

In parallel with the recovery of the power system, various activities were performed on the wet IT equipment (like cleaning and drying of disks, servers and switches). A good thing has been discovered that wet disks work, at least for a while, after they had been cleaned and dried.

As a strategy, we decided to replace damaged disks only for systems still under support in 2018, and to use spare parts we had in house, to recover, at least partially, the oldest systems still in production. It has been an interlocking game! Regardless massive disk failure and delayed intervention from support service, we have managed to recover 2.6 PB (out of 7.8 PB) of scientific data stored on OS6800v3 system. For more details see [6].

6 Out of the mud

In January, after chillers were restarted, we could proceed to re-open all the services, including part of the farm (at the beginning only 50,000 HS06, 1/5 of the total power capacity) and

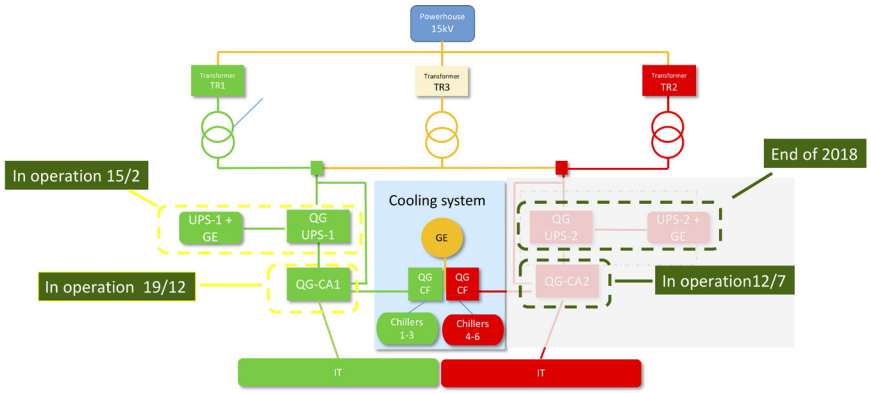

Figure 8. The Power Center recovery step by step

progressively the storage systems. The first experiments to resume operations at CNAF were ALICE, Virgo, Darkside[7]: the storage system used by Virgo and Darkside had been easily recovered after Christmas break, while ALICE is able to use computing resources relaying on remote storage. In February and March, we were able to progressively re-open the services for all the experiments according to the recovery plan for the storage systems (see Section 5).

We avoided part of the farm recovery thanks to the setup of a partition of the farm at CINECA[8] super-computing center (216 WNs for a total computing power of of $\sim$ 180 kHS06): a dedicated fiber directly connecting INFN Tier-1 core switches to our aggregation router at CINECA was installed and, via a couple of Infinera DCI, a link with a bandwidth of 500 Gbps (upgradable to 1.2 Tbps) was configured. Due the low latency (the RTT is 0.48 ms vs. 0.28 ms measured on the CNAF LAN), there is no need of disk cache and the WNs directly access the storage located at CNAF; in fact, the efficiency of the jobs is comparable to the one measured on the farm partition at CNAF.

7 Lessons learned

Designing our data center, we thought to have foreseen all possible incidents (e.g. fires, power cuts...): this was not true (and even possible). The only threat from water was supposed to come via intense raining. Indeed, waterproof doors were installed some years ago (after a heavy rain) and no evidence of other issues were found. So, a revision of the impermeabilization of the data center is definitely needed and ongoing.

An interesting aspect we have learned is that wet disks and tapes are not definitely lost. After having been carefully cleaned and dried, disks can be powered on for the required time to copy data out of them. For tapes, the "trick" is to reread critical points several times (and transferring data to a new media)[1].

A quite obvious lesson is that no experiment should base its computing on a single site or, even worse, store the data in a single place. This is precisely what happened for some smaller experiments using only CNAF as computing and data management facility.

The most important lesson is how fundamental it is to have a skilled and motivated staff: such a quick recovery would not have been possible without the huge and prolonged effort of all CNAF people.

[1] A special drive (i.e. one ignoring read errors!) is needed.

8 The future

Even before the flooding we were looking for a new location for our Tier-1, given the limited expandability of our data center. In fact, according to our estimates, it will only be able to host resources up to 2023. Considering the foreseen needs for HL-LHC (including the scenario of the so-called data-lake model) and the further expansions due to the astro-particle experiments, the plan was to build a data center with an IT power of up to 10 MW. This has become more urgent after the flooding.

An opportunity is given by the new ECMWF (European Centre for Medium-Range Weather Forecasts)[9] center which will be hosted in Bologna, in a new Technopole area, starting from 2019[10]. In the same area the INFN Tier-1 and the CINECA computing centers can be hosted too: funding has been guaranteed to INFN and CINECA by the Italian Government for this. The goal is to have the new data center for the INFN Tier-1 fully operational by the end of 2021.

Acknowledgement

Immediately after the accident, it was not clear whether the return to normal state would have been possible. The rapid recovery of the data center and the return to full operation are due to the excellent work and dedication of our staff and the support of INFN management.

References

[1] Italian Research and Academic Network (http://www.garr.it/en/)
[2] https://www.slac.stanford.edu/BFROOT/
[3] https://www-cdf.fnal.gov/
[4] http://www.virgo-gw.eu/
[5] Research and Competitiveness project (https://www.recas-bari.it/index.php/en/).
[6] L. dell'Agnello, "The flood", "INFN-CNAF Annual Report 2017", pp. 154-162 (2018) - ISSN 2283-5490 (online)
[7] http://darkside.lngs.infn.it/
[8] https://www.cineca.it/en/
[9] https://www.ecmwf.int/
[10] https://www.ecmwf.int/en/about/media-centre/news/2017/ecmwfs-new-data-centre-be-located-bologna-italy-2019