INFN CNAF Data Center

Luca dell'Agnello

CTSC review

Bologna, October 9 2020

CNAF Data Center

- It is the INFN main computing center
- Scientific computing
  - Tier1 for the four WLCG experiments and Tier-2 for LHCb
    - Resources for a total of ~50 scientific collaborations (July 2020)
  - Two small HPC clusters
  - General-purpose cloud infrastructure

~405 kHS06 for the computing farm
Before the end of 2020: + ~140 kHS06
~40 PB of disk
Before the end of 2020: + 8 PB of disk
~90 PB of tapes.

- ICT computing infrastructure
  - Information System
  - National ICT Services

Personnel

- Personnel situation improved
  - 1 person added in farming and 1 person in networking group
- Changed leader of Facility Management group
  - Former coordinator on leave

The Tier1: highlights (1/2)

- Migration from LSF to HTCondor accomplished
  - Finalized for the main experiments end of March 2020
  - Completed in July to allow some users to complete the validation of the new system.
- Effort related to Covid-19
  - In the first half of March 2020, nearly half of the HTC farm has been dedicated to biomedical simulations in collaboration with Sibylla biotech
  - Participation to EU project Exscalate4Cov (set-up if a repository to store output of simulations completed on CINECA HPC system)
  - WLCG experiments use part of their pledge for F@H jobs (completely transparent to us)
- Tests on CINECA HPC system
  - Part of Marconi A2 is being used, in a transparent way, from WLCG experiments
  - Just starting now tests with Marconi 100 (newest HPC system)
- Installation and validation of the new tape library

The Tier1: highlights (2/2)

- New configuration of CNAF cloud instance completed.
  - Cloud@CNAF now including both SDDS and Tier1 resources.
- Complete review of monitoring and accounting system.
- Comprehensive set of KPIs defined
- Adoption of the issue system maintained by the ICT National Services team
  - All internal tasks are tracked
- Installation and validation of new edge router
- Incident on main power line (August 2019)
- One of the KS systems completely refurbished and out into operation (January 2020)
  - Extraordinary maintenance required after 10 years of operations

Migration from LSF to HtCondor

- Process started in the 2nd half of 2018 and completed beginning of Q3 2020
  - Largest part of farm migrated in March 2020
  - Last fraction (~10 kHS06) migrated end of June 2020 to allow small experiments to test the new system
- Also Cream CEs migrated to Condor CEs
- Fruitful support from Condor Team
- Building a INFN knowledge base

Integration with CINECA HPC system Marconi A2 (1/4)

CINECA

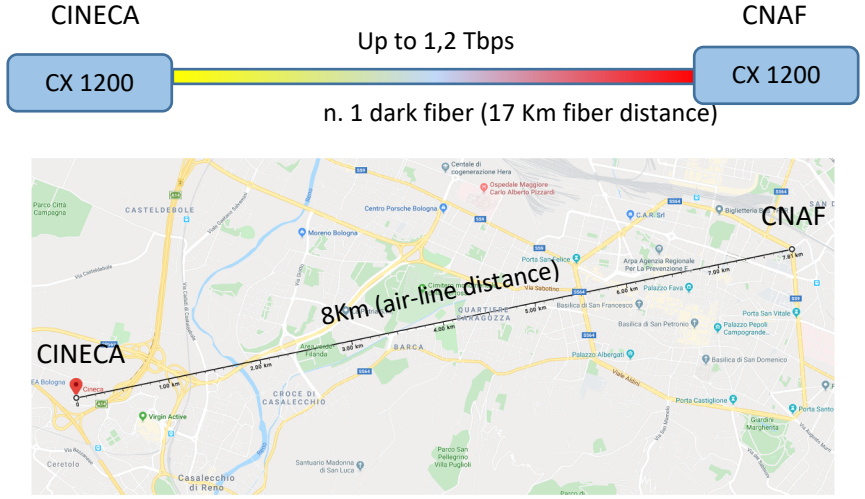
Marconi A2 Partition

3600 nodes with 1 Xeon Phi 2750 (KNL) at 1.4 GHz and 96 GB of RAM

68 cores/node, 244800 cores

Peak Performance: ~11 Pflop/s

CINECA – CNAF DCI



"THE GRANT"

The LHC Italy community successfully applied for a **"PRACE Project Access"** on the CINECA KNL partition

30McoreH allocated for running LHC Jobs

CNAF

Standard "GRID-like" HTC farm (30k cores, 400 kHS06)

•41 PB of disk

•90 PB of tapes on 2 libraries

Integration with CINECA HPC system Marconi A2 (2/4)

Matching LHC workloads with HPCs is not that easy because they derive from different "User requirements".

In the HPC centers there are usually **strict site policies not matching the LHC user requirements**

- Ad-hoc operating system, limited/absent external connectivity, user policies only for individuals, node hardware setup, ...

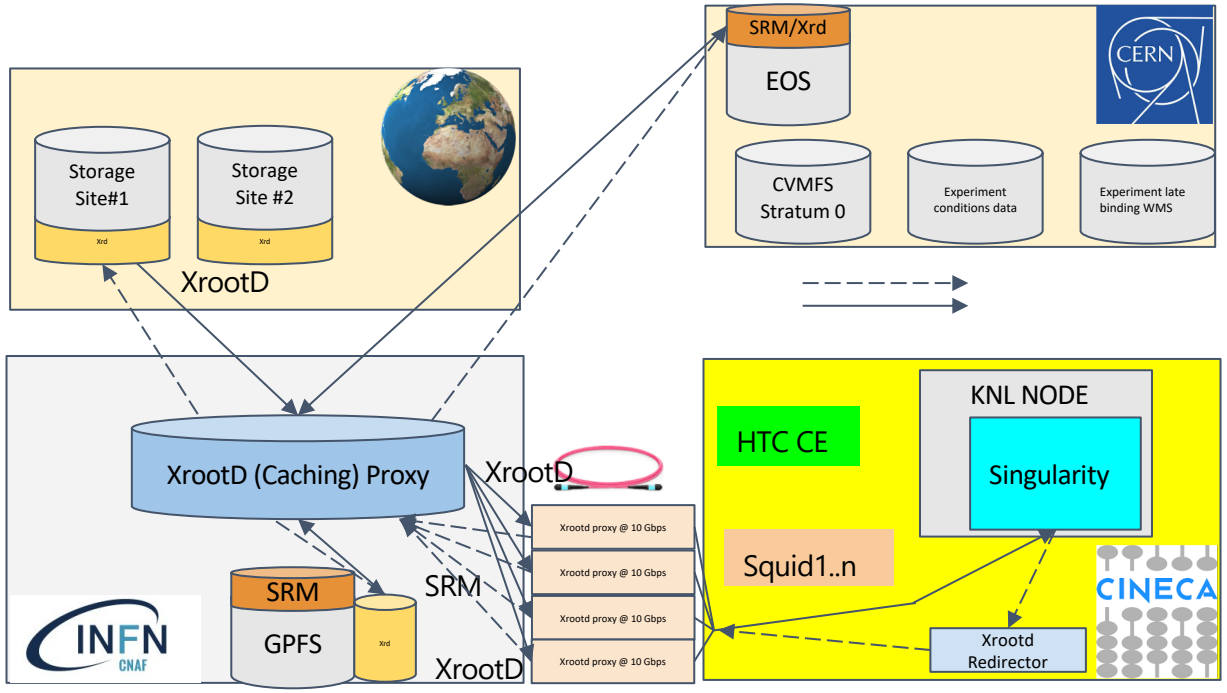
A lot of work needed to establish a "trust model" With CINECA and to understand the peculiarity of the HPC platforms running at CINECA and to adapt each workflows to run together.

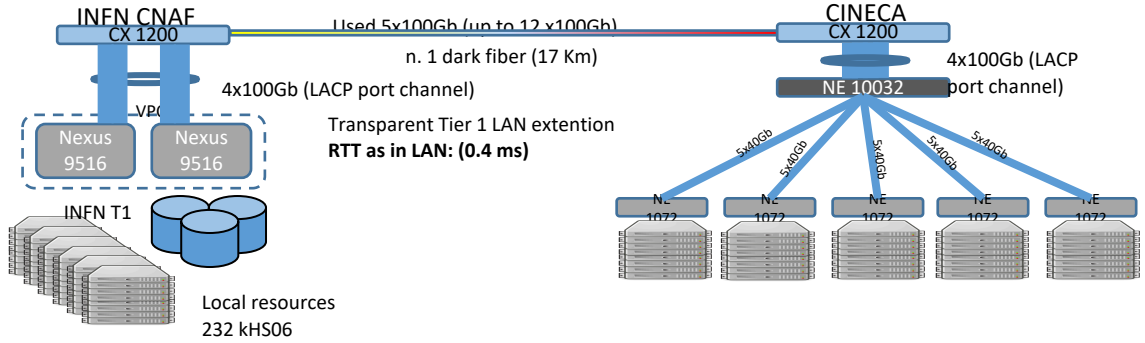
Many people involved but very enriching experience for everyone.

A standard CINECA Marconi A2 is node configured with	A typical WLCG node has
An Intel(R) Xeon Phi(TM) CPU 7250 @ 1.40GHz: 68 or 272(HT4x) cores, x86_64, rated at ~¼ the HS06 of a typical Xeon per core	1-2 Xeon-level x86_64 CPUs: typically 32-128 cores, O(10 HS06/thread) with HT on
96 GB RAM, with ~10 to be reserved for the OS: 1.3-0.3 GB/thread	2GB/thread, even if setups with 3 or 4 are more and more typical (so a total 64-256 GB)
No outgoing connectivity from the node (only a very limited access to the external networks via NAT on bastion hosts) – CINECA overall WAN connectivity limited to 10Gbps (more than enough for its core users workflows)	Full outgoing external connectivity, with sw accessed via CVMFS mounts; additional experiment specific access needed (condition DBs, input files via remote Xrootd, ...)
No local disk (large scratch areas via GPFS/Omnipath)	O(20 GB/thread) local scratch space
Access to batch nodes via SLURM; Only Whole Nodes can be provisioned, with 24 h lease time	Access via a CE. Single thread and 8 thread slots are the most typical; 48+ hours lease time
Access granted to individuals (via passport / fiscal code identification)	Access via pilots and late binding; VOMS AAI for end-user access

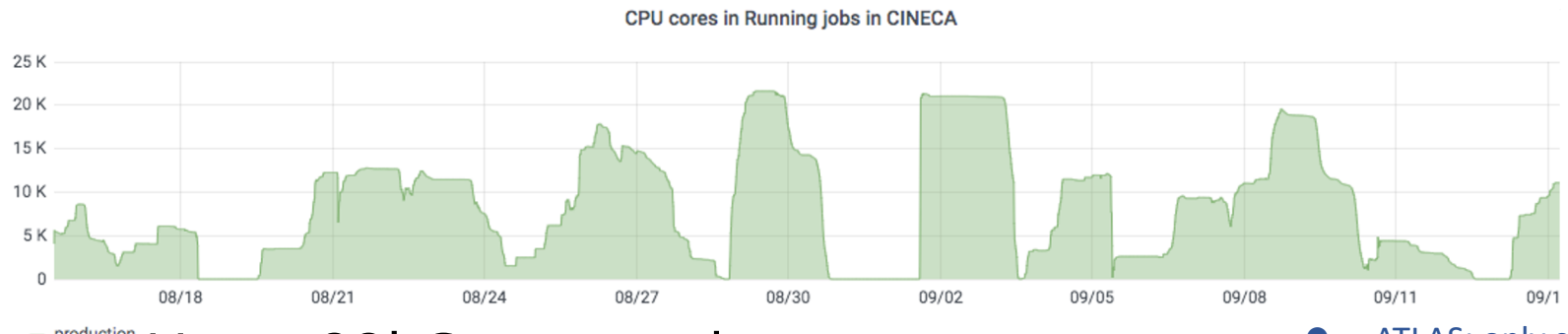
Integration with CINECA HPC system Marconi A2 (3/4)

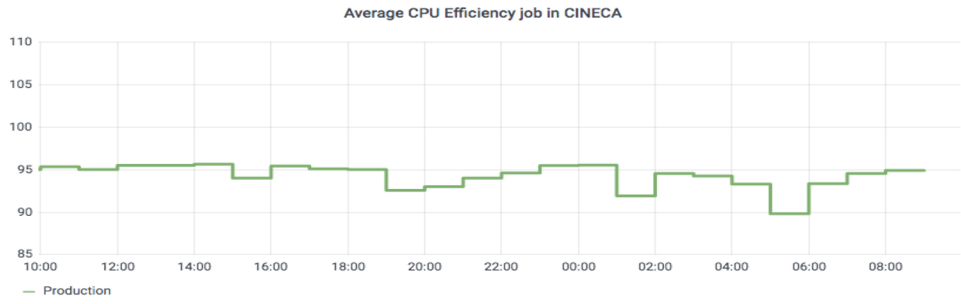
- Ad hoc Condor-CE installed @CINECA as interface with Slurm (+4 squids)
  - Marconi A2 seen as part of INFN Tier-1 in transparent way for users (as much as possible)
- Payloads download from CERN via CINECA GPN (10 Gbps)
- Data (XrootD only) traffic routed via 4 gateways (4x10 Gbps) through DCI extention between CNAF and CINECA





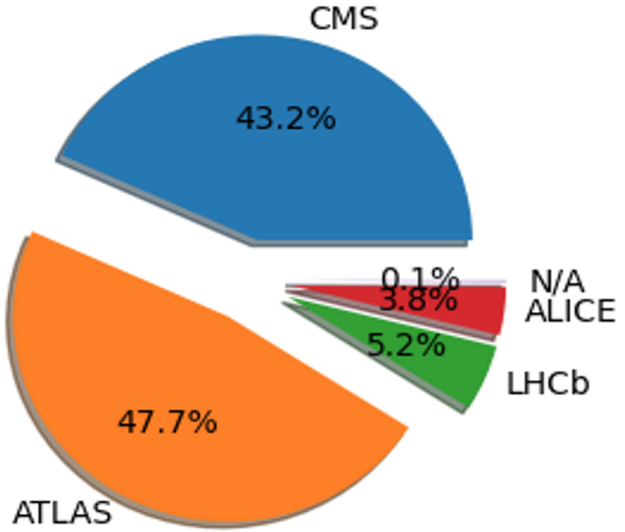
Integration with CINECA HPC system Marconi A2 (4/4)

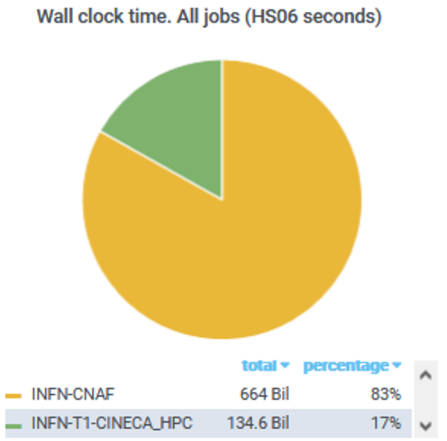




- Up to 22kCores used
- Success rates 85-90% (mostly timeouts and application errors)
- CPU efficiency ~95% (with data intensive jobs!)
- Utilization of the link very low (only a small fraction of the CMS WFs require sizeable I/O in this period)
- **70 Million Core hours used to date**

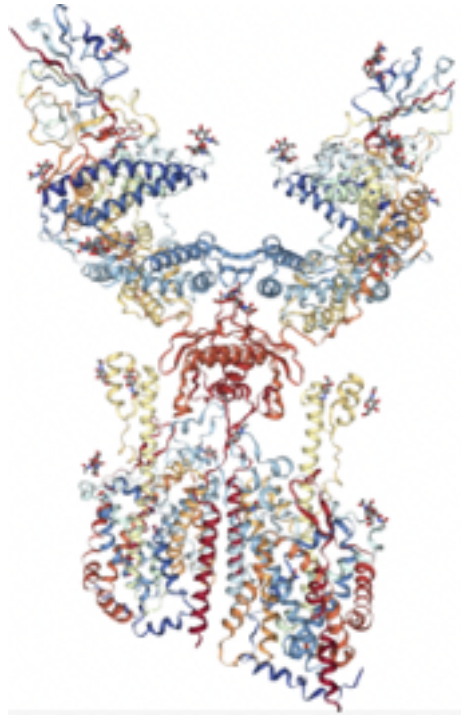
- ATLAS: only simulation jobs, using 48 threads per node
- LHCb: simulation jobs, up to 136 threads per node
- ALICE: do not use WMS, run local Run-3 O2 tests
- CMS: all production jobs, up to 128 threads per node





Sibylla Biotech & INFN against COVID-19 (1/2)

- Sibylla Biotech is a Biotechnology startup created in late 2019
  - Academic Stakeholders: INFN, Univ Trento, Univ Perugia
  - In the staff, INFN associates and former staff (Theoretical Physicists, Experimental Physicists)
  - General idea: **port HEP-th derived ideas to the simulation of molecular dynamics & drugs design**; method claimed to be 10000x faster in following the dynamics of folding
- General strategy: **Pharmacological Protein Inactivation by Folding Intermediate Targeting (PPI-FIT)**
  - interfere with the final shape of the proteins in human body targeted as attachment by the virus. In this way disable the capability of the virus to link to human targets.
  - In this specific case: simulate the folding of the protein ACE2 (angiotensin-converting enzyme 2), the pulmonary receptor that binds to the spikes of Covid
- **Goal: O(1000) simulations** with different initial states
  - **~32000 cores for 10 days needed**



Sibylla Biotech & INFN against COVID-19 (2/2)

- Joint effort of Tier-1 and several Tier-2s
- Dedicated queue and storage area created and tested over week-end
  - Computing requirements fitting quite well within Tier-1 environment (no experiment specific solutions)
    - **AVX-2** vectorization (no Opterons, …)
    - **32+ threads/job**
    - **GCC 9.x** (for AVX-2 optimal support) **on CC7**
      - **WNs with Centos7 (for CC7) and CVMFS to distribute GCC 9.x**
    - Strongly checkpointed executable, with a structure muxing inputs and outputs
      - Common shared work area between the nodes **(e.g. GPFS)**
- In the end, job length was smaller than 10 days (~ 7 on average)
- 1200 jobs processed out of 1000 base request
- CNAF started first, and in some case could perform 2 round of jobs
- 2 intermediate "blocking targets" found
- 35 existing and approved for human treatment molecules identified as able to interfere with the creation of the bounding site identified

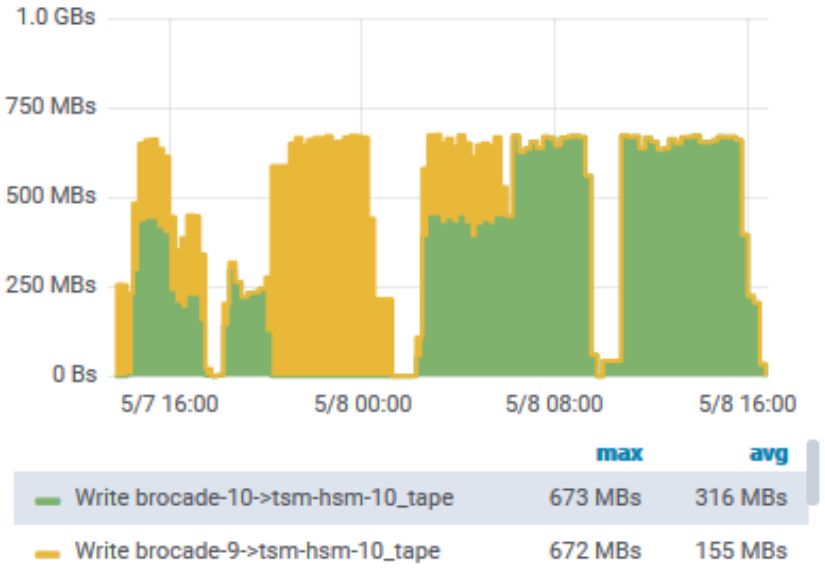
SITE	Max number of cores for Covid Sibylla simulation	CPU source	Core h dedicated to the project
CNAF	17500	Generic	3630067
PI	2556	CMS, Theory	498284
BA	1280	ReCaS, CMS, ALICE	251725
RM1	1536	ATLAS	347525
MI	1280	ATLAS	145103
LNL-PD	2560	CMS, ALICE (70%/30%)	416083
NA	1024	ReCaS, ATLAS	185536
LNF	1280	ATLAS	128576
TOT CPU Cores	**29016**	**TOT Core.h**	**5602899**

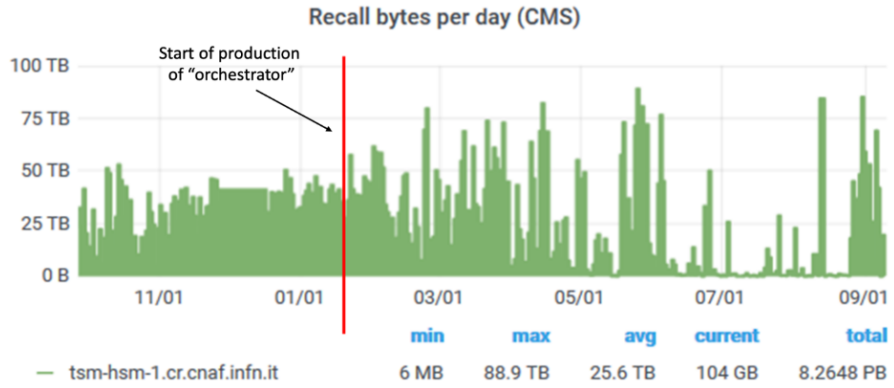
The Committee believes that CNAF should review the architecture and the procurement models to verify that the current tape-based strategy on enterprise grade drives and cartridges is still the most effective way for archival. In particular, the usage of LTO (Linear Tape Open) media should be scrutinized.

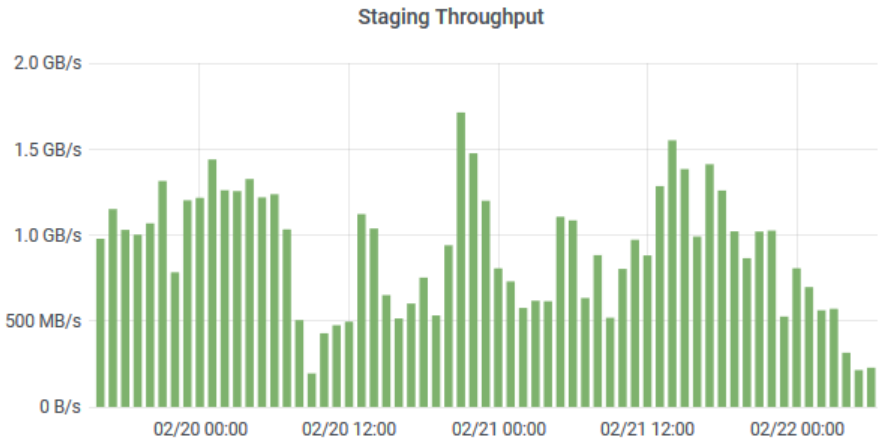
- Moving from Oracle to IBM technology, we had a sharp reduction of the cost of the media (~40%)
- LTO technology was not considered for the new library mainly for contingent issues related to legal dispute between tape producers
- In the medium term, we will explore the possibility to convert the old library to LTO technology
- As an alternative we could buy a new library (based on LTO) to replace the first one
  - The cost of support is one of the driving factors to be considered for this choice
  - Need to factorize the cost of maintenance of drives respect to library (first indication: ~25 kE/year)

Tape library (1/2)

- IBM TS4500 in production since January 2020
  - 19 last generation enterprise tape drives (TS11600)
    - 20 TB/tape
    - 400 MB/s both for reading and writing
- Overall data rate capacity (7.6 + 4 GB/s) to and from tape.
- Optimization of access to tapes through an orchestrator for dynamic allocation of drives to experiments.







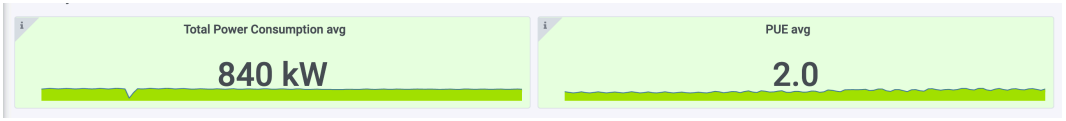
Tape library (2/2)

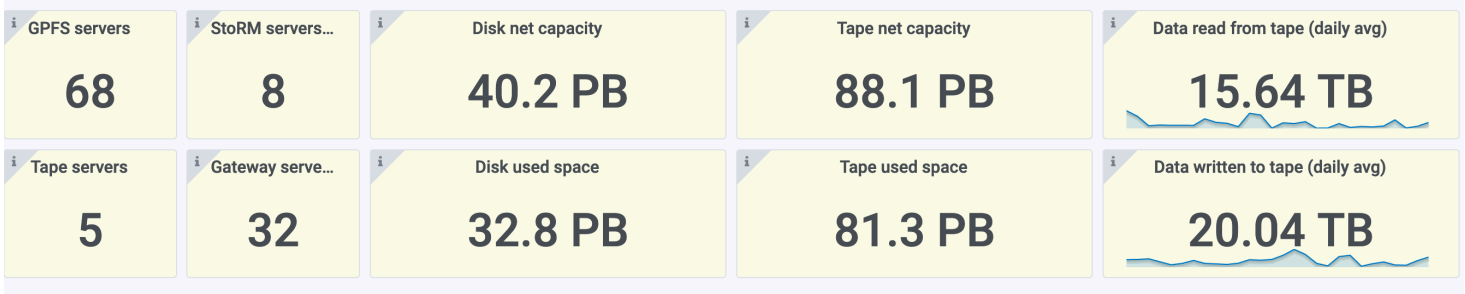
- Moving away from Oracle to IBM technology we have noted a sharp reduction of the cost of the media (~40%).
- In the medium term, we will explore the possibility to convert the old library to LTO technology
- As an alternative we could buy a new library (based on LTO) to replace the first one
  - The cost of support is one of the driving factors to be considered for this choice
    - Need to factorize the cost of maintenance of drives respect to library (first indication: ~25 kE/year)

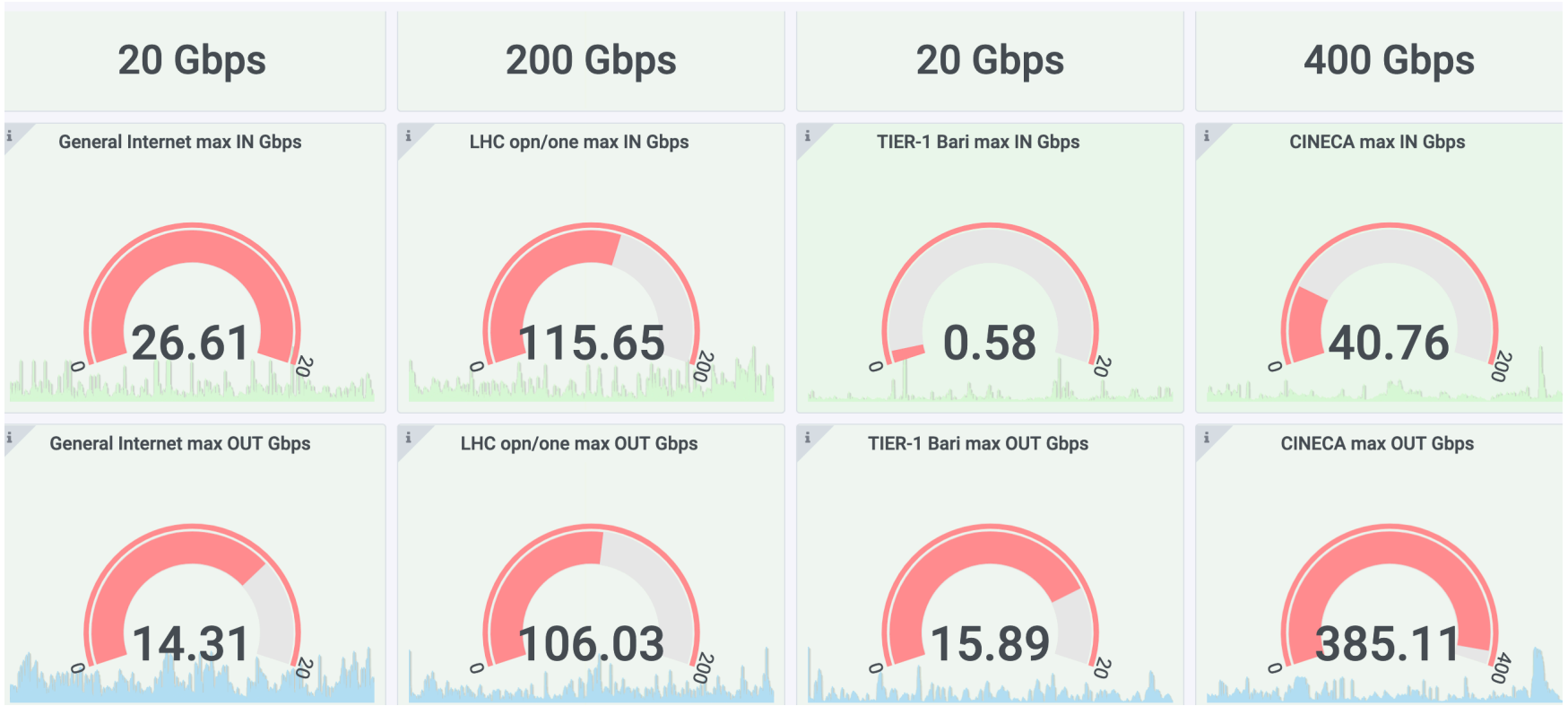
- The monitoring and alarm systems, based on Graphana and Sensu, have been completely reviewed.
- Two different views (https://t1metria.cr.cnaf.infn.it/)
  - Private - for data center administrators only (all IT systems and services and various infrastructure metrics are covered)
  - Public - use of the resources from the point of view of experiments.

The Committee believes that a systematic and richer set of KPIs will yield a gain in quality and efficiency. The Committee recommends that CNAF embarks on a study towards achieving this goal.

- An extensive set of KPIs has been defined for all aspects of the data center. A subset is presented at the same page of the monitoring (https://t1metria.cr.cnaf.infn.it/)
- We are reconsidering all these in order to reduce to those really meaningful to give the performances and state of the data center






20 Gbps
200 Gbps
20 Gbps
400 Gbps
General Internet max IN Gbps
26.61
0
20
LHC opn/one max IN Gbps
115.65
0
200
TIER-1 Bari max IN Gbps
0.58
0
20
CINECA max IN Gbps
40.76
0
200
General Internet max OUT Gbps
14.31
0
20
LHC opn/one max OUT Gbps
106.03
0
200
TIER-1 Bari max OUT Gbps
15.89
0
20
CINECA max OUT Gbps
385.11
0
400

With TCO in mind, CNAF should reconsider the continuation of maintenance support contract in favour of purchasing additional storage that could compensate and exceed the resource reduction due to hardware failures over time.

- The cost of the maintenance contract is not strictly proportional to the total price: in the last tender procedures we have recorded an impact, due to maintenance, ranging from ~4% to ~11% of the total cost.

Multi-year framework agreement

Year	€/TB-N	Maint. cost (€)	Impact of maint. (%)
2016	123,57	57475	7.8
2017	86,87	78330	11,3
2018	104,15	94905	7,9
2019	114,66	25000	4,3

- Framework agreement not always viable solution
  - Budget fixed on yearly basis
  - currently testing alternative solutions to our current storage model, making it difficult to adopt a technological choice for more than one tender.
- Working group, within the framework of the collaboration agreement with CERN-IT., In order to evaluate the TCO of storage solutions

- Preliminary tests have been performed on EOS (June-July 2019) and CEPH (September 2019-January 2020)
- The hardware for both testbeds was recycled from old storage systems
  - not suitable for these technologies (the storage system standard at CNAF is composed of a relatively limited number of servers), thus not allowing us to draw definitive conclusions.
- Planning to repeat these tests extensively in the framework of the collaboration with CERN-IT
  - interest in verifying the viability of a system based on both CEPH (for the disk management part) and EOS (for the services).
  - A dedicated testbed with 4 PB, 8 servers and 8 JBODs is already on-line and available for tests.
- A significant part of the 2020 storage acquisition (about 4 PB out of 15) as an EOS/CEPH pilot system to be used in production for one of the experiments running at CNAF.
- CLOUD@CNAF infrastructure is equipped with a CEPH system, which has recently seen the addition of 8 disk servers (with 24 16 -B disks inside – no JBODs used in this case) for a total of 3 PB.

- CPU resources are organized in a general-purpose farm
  - each experiment has access to a share according to the agreed pledge
  - fair-share mechanism implemented (dynamic redistribution of unused resources)
    - This works better for WLCG experiments (constant "pressure" on batch systems)
    - Undergoing discussion with the stakeholders of astro-particle experiments in order to group all their resources in a unique pool, thus allowing an even better redistribution
  - Currently, manual intervention is required only to satisfy requests of additional resources to be extracted from the common pool: since the amount of resources beyond the pledges is very limited it is unfeasible to allow a completely automatic mechanism besides the fair-share.
- For storage, the main difficulty is that very seldom used disk is freed (unless to store other data by the same collaboration): hence a true dynamic system is unlikely.
- Anyway, besides the 'big players' (i.e. LHC experiments and a few others), storage is allocated to experiments with a common file-system, allowing a good optimization and even "thin-provisioning"

- In preparation for the migration to the Tecnopolo, two complementary initiatives were launched: a project for the design, both in terms of hardware and services, the new data center ("CNAF reloaded") and a collaboration with CERN-IT on specific topics, among which: storage models, network architecture, container technology. These two projects will be the main activity of CNAF in the next year.

- Funding for computing resources of experiments is negotiated between each collaboration and a referral committee (external to CNAF)
- On the other hand, the analysis of the experiments' requirements in terms of needed performances (e.g. Quality of Service for the storage, needed bandwidth, etc.) is done before the actual allocation of the pledges, with the collaboration of User Support and of the CNAF technical groups.
- Unless explicitly requested (and funded) LTDP is meant as bit preservation only
  - Only 4 experiments have been ceased the production phase
  - For CDF we have the complete framework for analysis up and running (and used!)

User support at CNAF is a quite challenging activity due to the large variety of supported projects with different requirements. In addition, the user support personnel is mainly temporary postdocs, with frequent rotation, making a difficult task to keep the knowledge base.

- User support is a team led by a senior technologist and composed of from 4 to 6 postdocs, usually physicists.
  - first level of support passing to the other Tier1 teams (in particular Storage, Farming and Network) the cases involving a deeper technical knowledge.
  - They are an interface to the experiments, understanding their needs, providing advisory to design their computing model, helping the experiments to fix their problem with data transfer, data storage and data processing.
  - This approach, based on the supervision of a person with deep experience and a young team, has worked quite well in the years and it was effective in solving problems.
- From the other side it is true that there is a high turnover rate, that create issues in the continuity of the support. However, to mitigate this situation, more experienced supporters, now moved to others Tier1 operations teams (Farming or Storage) still contribute for a fraction of their working time to the user support activities and to the training of the hired postdocs.

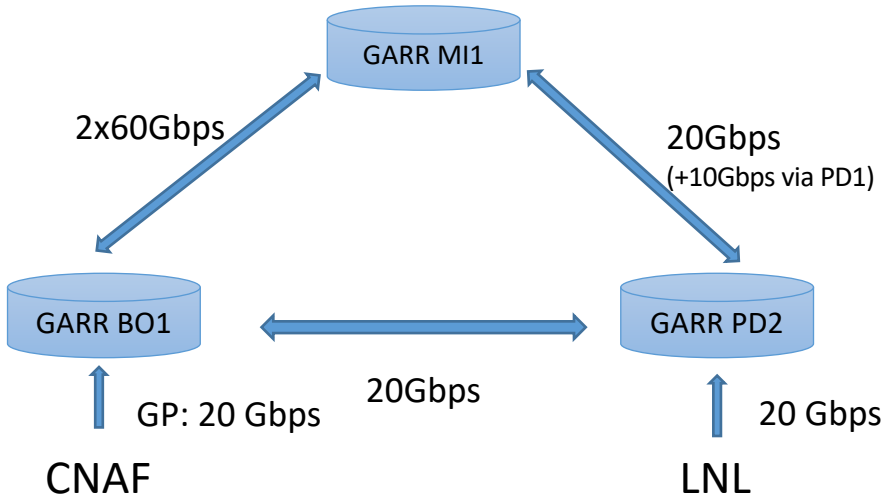
National ICT Services

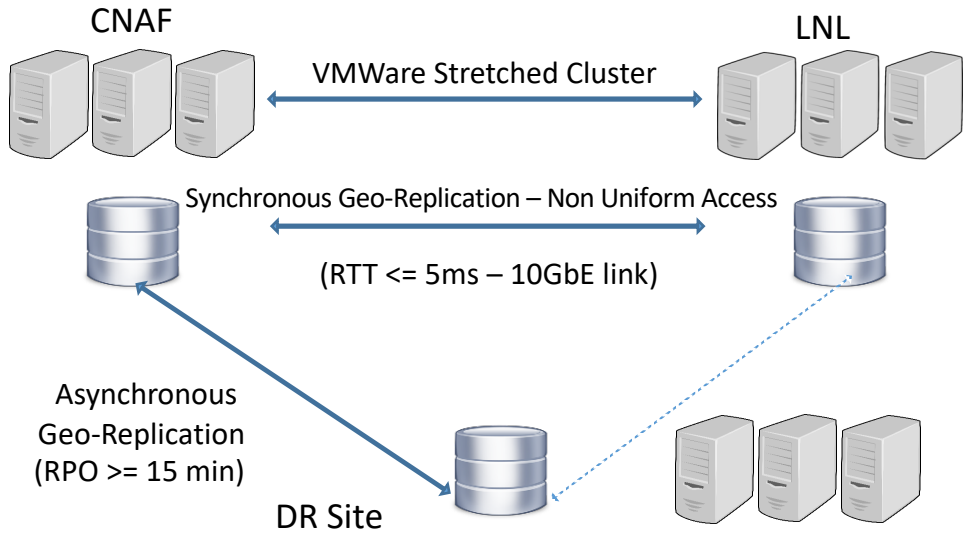
CTSC Mid-Term Review

Business Continuity Infrastructure

Assure service availability even in the event that CNAF has some major failure

- HA provided by the infrastructure
- No need to modify applications
- Recovery Point Objective ~ 0
- Recovery Time Objective -> 0





Half infrastructure deployed at Laboratori Nazionali di Legnaro, taking into account

- Distance > 100Km with an RTT ~ 2,2ms
- Tier2 data center already in place (cooling and electrical continuity)
- Direct connection to CNAF plus an alternative path for General Internet

Business Continuity Infrastructure

- New Infrastructure in production from July 4th, 2019
- Systems available for all CCR groups working on National ICT Services (INFN Information System, Mailing, AAI, OKD, etc.)
- 205 VM running at present

The infrastructure is working pretty well, both in terms of performances and resiliency.

Upgrade plan:

- 2020: Doubled (or a bit more) computing power. Added dedicated disk pool for NAS access with synchronous geo-replication and local caching
- 2021: Planned computing upgrade by adding another 50% of the current resource capacity

Coutinuos scouting for technologies and infrastructures able to satisfy service requirements

INFN-Cloud will be exploited for non critical services

Incident

- August 6th, 2019, at 17.30 CET, the protected power line of storage and network systems failed due to short circuit
- The busway connecting the dynamic UPS (or KS) to the main electrical panel was subjected to a short circuit and, therefore, a mechanical joint of the KS broke. After a second, the bypass of the KS was activated, and the power was restored.
- The first inspection revealed the fault of the KS system only; in a second moment (August 10), also the problem on the busway was evident.
- The down was closed on August 21st at 20.00 CET
- The cause of the problem was the degrade insulation on external part of the busway (they are regularly checked)
- After the incident, all busways have been replaced

Timeline of the incident

- 2019-08-06 17.45 CET – Start of the incident (power interruption)
- 2019-08-06 19.00 CET – Inspection of the system: fault of the KS is revealed but no evidence of problems on the distribution system. Support of the KS system is alerted.
- 2019-08-08 09.00 CET – KS support intervention; problem escalated to KS producer.
- 2019-08-10 21.00 CET – Mechanical joint of the dynamic KS restored.
- 2019-08-10 21.30 CET – Problem on the busway discovered
- 2019-08-12 15:00 CET – Substitution of faulty parts of the busway started
- 2019-08-14 13:00 CET – Busway restored
- 2019-08-19 13:00 CET – Installation of a static UPS on the second line started.
- 2019-08-21 15:00 CET – Installation of the static UPS on the second line completed and validated
- 2019-08-21 17:00 CET – All IT services restored
- 2019-08-21 19:00 CET – All IT services checked and validated
- 2019-08-21 20:00 CET – Down closed

GEMSS

- GEMSS (Grid Enabled Mass Storage System) is a full HSM solution
  - Thin software layer developed in-house
  - Integration of GPFS, TSM and StoRM (developed in-house too)
  - Allows a reordered recall of files from tapes (optimizing the tape mounts)
  - Manages migration from disk buffer to tape library
- 5 HSM nodes (1 for each LHC experiments and 1 for all the others) provide all the data movements between disks and tapes
  - Interconnection via TAN (16 Gbps FC based)
  - 1 cold-spare systems available
- Advantages: high reliability and low effort needed for its operation
- Interest shown by IBM to include in the TSM suite